

Restricted Isometry Property for General p-Norms*

Zeyuan Allen-Zhu[†]

Rati Gelashvili[†]

Ilya Razenshteyn[†]

February 24, 2015

Abstract

The Restricted Isometry Property (RIP) is a fundamental property of a matrix which enables sparse recovery. Informally, an $m \times n$ matrix satisfies RIP of order k for the ℓ_p norm, if $\|Ax\|_p \approx \|x\|_p$ for every vector x with at most k non-zero coordinates.

For every $1 \leq p < \infty$ we obtain almost tight bounds on the minimum number of rows m necessary for the RIP property to hold. Prior to this work, only the cases $p = 1$, $1 + 1/\log k$, and 2 were studied. Interestingly, our results show that the case $p = 2$ is a “singularity” point: the optimal number of rows m is $\tilde{\Theta}(k^p)$ for all $p \in [1, \infty) \setminus \{2\}$, as opposed to $\tilde{\Theta}(k)$ for $k = 2$.

We also obtain almost tight bounds for the column sparsity of RIP matrices and discuss implications of our results for the Stable Sparse Recovery problem.

1 Introduction

The main object of our interest is a matrix with *Restricted Isometry Property for the ℓ_p norm* (RIP- p). Informally speaking, we are interested in a linear map from \mathbb{R}^n to \mathbb{R}^m with $m \ll n$ that approximately preserves ℓ_p norms for *all* vectors that have only few non-zero coordinates.

More precisely, an $m \times n$ matrix $A \in \mathbb{R}^{m \times n}$ is said to have (k, D) -RIP- p property for sparsity $k \in [n] \stackrel{\text{def}}{=} \{1, \dots, n\}$, distortion $D > 1$, and the ℓ_p norm for $p \in [1, \infty)$, if for every vector $x \in \mathbb{R}^n$ with at most k non-zero coordinates one has

$$\|x\|_p \leq \|Ax\|_p \leq D \cdot \|x\|_p .$$

In this work we investigate the following question: given $p \in [1, \infty)$, $n \in \mathbb{N}$, $k \in [n]$, and $D > 1$,

What is the smallest $m \in \mathbb{N}$ so that there exists a (k, D) -RIP- p matrix $A \in \mathbb{R}^{m \times n}$?

Besides that, the following question arises naturally from the complexity of computing Ax :

What is the smallest column sparsity d for such a (k, D) -RIP- p matrix $A \in \mathbb{R}^{m \times n}$?

(Above, we denote by column sparsity the maximum number of non-zero entries in a column of A .)

* An extended abstract of this paper is to appear at the 31st International Symposium on Computational Geometry - SoCG 2015.

[†]MIT CSAIL, email: zeyuan@csail.mit.edu, {gelash,ilyaraz}@mit.edu

1.1 Motivation

Why are RIP matrices important? RIP-2 matrices were introduced by Candès and Tao [CT05] for decoding a vector f from corrupted linear measurements $Bf + e$ under the assumption that the vector of errors e is sufficiently sparse (has only few non-zero entries). Later Candès, Romberg and Tao [CRT06] used RIP-2 matrices to solve *the (Noisy) Stable Sparse Recovery* problem, which has since found numerous applications in areas such as compressive sensing of signals [CRT06, Don06], genetic data analysis [KBG⁺10], and data stream algorithms [Mut05, GI10].

The (noisy) stable sparse recovery problem is defined as follows. The input signal $x \in \mathbb{R}^n$ is assumed to be close to k -sparse, that is, to have most of the “mass” concentrated on k coordinates. The goal is to design a set of m linear measurements that can be represented as a single $m \times n$ matrix A such that, given a *noisy sketch* $y = Ax + e \in \mathbb{R}^m$, where $e \in \mathbb{R}^m$ is a noise vector, one can “approximately” recover x . Formally, the recovered vector $\hat{x} \in \mathbb{R}^n$ is required to satisfy

$$\|x - \hat{x}\|_p \leq C_1 \min_{k\text{-sparse } x^*} \|x - x^*\|_1 + C_2 \cdot \|e\|_p \quad (1.1)$$

for some $C_1, C_2 > 0$, $p \in [1, \infty)$, and $k \in [n]$.

(In order for (1.1) to be meaningful, we also require $\|A\|_p \leq 1$ —or equivalently, $\|Ax\|_p \leq \|x\|_p$ for all x — since otherwise, by scaling A up, the noise vector e will become negligible.)

We refer to (1.1) as the ℓ_p/ℓ_1 *guarantee*. The parameters of interest include: the number of measurements m , the column sparsity of the measurement matrix A , the approximation factors C_1 , C_2 and the complexity of the recovery procedure.

Candès, Romberg and Tao [CRT06] proved that if A is $(O(k), 1 + \varepsilon)$ -RIP-2 for a sufficiently small $\varepsilon > 0$, then one can achieve the ℓ_2/ℓ_1 guarantee with $C_1 = O(k^{-1/2})$ and $C_2 = O(1)$ in polynomial time.

The $p = 1$ case was first studied by Berinde *et al.* [BGI⁺08]. They prove that if A is $(O(k), 1 + \varepsilon)$ -RIP-1 for a sufficiently small $\varepsilon > 0$ and has a certain additional property, then one can achieve the ℓ_1/ℓ_1 guarantee with $C_1 = O(1)$, $C_2 = O(1)$.

We note that *any* matrix A that allows the (noisy) stable sparse recovery with the ℓ_p/ℓ_1 guarantee *must have the* (k, C_2) -RIP- p *property*. For the proof see Appendix A.

Known constructions and limitations. Candès and Tao [CT05] proved that for every $\varepsilon > 0$, a matrix with $m = O(k \log(n/k)/\varepsilon^2)$ rows and n columns whose entries are sampled from i.i.d. Gaussians is $(k, 1 + \varepsilon)$ -RIP-2 with high probability. Later, a simpler proof of the same result was discovered by Baraniuk *et al.* [BDDW08]¹. Berinde *et al.* [BGI⁺08] showed that a (scaled) *random sparse binary matrix* with $m = O(k \log(n/k)/\varepsilon^2)$ rows is $(k, 1 + \varepsilon)$ -RIP-1 with high probability².

Since the number of measurements is very important in practice, it is natural to ask, how optimal is the dimension bound $m = O(k \log(n/k))$ that the above constructions achieve? The results of Do Ba *et al.* [DIPW10] and Candès [Can08] imply the lower bound $m = \Omega(k \log(n/k))$ for $(k, 1 + \varepsilon)$ -RIP- p matrices for $p \in \{1, 2\}$, provided that $\varepsilon > 0$ is sufficiently small.

Another important parameter of a measurement matrix A is its *column sparsity*: the maximum number of non-zero entries in a single column of A . If A has column sparsity d , then we can perform multiplication $x \mapsto Ax$ in time $O(nd)$ as opposed to the naive $O(nm)$ bound. Moreover, for sparse matrices A , one can maintain the sketch $y = Ax$ very efficiently if we update x . Namely, if we set $x \leftarrow x + \alpha \cdot e_i$, where $\alpha \in \mathbb{R}$ and $e_i \in \mathbb{R}^n$ is a basis vector, then we can update y in time $O(d)$ instead of the naive bound $O(m)$.

¹This proof has an advantage that it works for any subgaussian random variables, such as random ± 1 's.

²In the same paper [BGI⁺08] it is observed that the same construction works for $p = 1 + 1/\log k$.

p	rows m	column sparsity d	references
1	$\Theta(k \log(n/k))$	$\Theta(\log(n/k))$	[BGI ⁺ 08, DIPW10, Nac10, IR13]
$1 + \frac{1}{\log k}$	$O(k \log(n/k))$	$O(\log(n/k))$	[BGI ⁺ 08]
(1, 2)	$\tilde{\Theta}(k^p)$	$\tilde{\Theta}(k^{p-1})$	this work
2	$\Theta(k \log(n/k))$	$\Theta(k \log(n/k))$	[CT05, CRT06, Can08, BDDW08], [DIPW10, Cha10, NN13]
(2, ∞)	$\tilde{\Theta}(k^p)$	$\tilde{\Theta}(k^{p-1})$	this work

Table 1: Prior and new bounds on RIP- p matrices

The aforementioned constructions of RIP matrices exhibit very different behavior with respect to column sparsity. RIP-2 matrices obtained from random Gaussian matrices are obviously dense, whereas the construction of RIP-1 matrices of Berinde *et al.* [BGI⁺08] gives very small column sparsity $d = O(\log(n/k)/\varepsilon)$. It is known that in both cases the bounds on column sparsity are essentially tight.

Indeed, Nelson and Nguyễn showed [NN13] that any non-trivial column sparsity is impossible for RIP-2 matrices unless m is much larger than $O(k \log(n/k))$. Nachin showed [Nac10] that any RIP-1 matrix with $O(k \log(n/k))$ rows must have column sparsity $\Omega(\log(n/k))$. Besides that, Indyk and Razenshteyn showed [IR13] that every RIP-1 matrix ‘must be sparse’: any RIP-1 matrix with $O(k \log(n/k))$ rows can be perturbed slightly and made $O(\log(n/k))$ -sparse.

Another notable difference between RIP-1 and RIP-2 matrices is the following. The construction of Berinde *et al.* [BGI⁺08] provides RIP-1 matrices with non-negative entries, whereas Chandar proved [Cha10] that any RIP-2 matrix with non-negative entries must have $m = \Omega(k^2)$ (and this was later improved to $m = \Omega(k^2 \log(n/k))$ [NN13, AGMS14]). In other words, negative signs are crucial in the construction of RIP-2 matrices but not for the RIP-1 case.

1.2 Our results

Motivated by these discrepancies between the optimal constructions for RIP- p matrices with $p \in \{1, 1 + \frac{1}{\log k}, 2\}$, we initiate the study of RIP- p matrices for the general $p \in [1, \infty)$.

Having in mind that the upper bound $m = O(k \log(n/k))$ holds for RIP- p matrices with $p \in \{1, 1 + \frac{1}{\log k}, 2\}$, it would be natural to conjecture that the same bound holds at least for every $p \in (1, 2)$. As we will see, surprisingly, this conjecture is very far from being true.

Also, knowing that the column sparsity $d = O(k \log(n/k))$ can be obtained for $p = 2$ while $d = O(\log(n/k))$ can be obtained for $p = 1$, it is interesting to “interpolate” these two bounds.

Besides the mathematical interest, a more “applied” reason to study RIP- p matrices for the general p is to get new guarantees for the stable sparse recovery. Indeed, we obtain new results in this direction.

Our Upper Bounds. On the positive side, for all $\varepsilon > 0$ and all $p \in (1, \infty)$, we construct $(k, 1 + \varepsilon)$ -RIP- p matrices with $m = \tilde{O}(k^p)$ rows. Here, we use the $\tilde{O}(\cdot)$ -notation to hide factors that depend on ε , p , and are polynomial in $\log n$. More precisely, we show that a (scaled) *random sparse 0/1 matrix* with $\tilde{O}(k^p)$ rows and column sparsity $\tilde{O}(k^{p-1})$ has the desired RIP property with high probability.

This construction essentially matches that of Berinde *et al.* [BGI⁺08] when p approaches 1. At the same time, when $p = 2$, our result matches known constructions of non-negative RIP-2 matrices

based on the incoherence argument.³

Our Lower bounds. Surprisingly, we show that, despite our upper bounds being suboptimal for $p = 2$, they are essentially tight for every constant $p \in (1, \infty)$ except 2. Namely, they are optimal both in terms of the dimension m and the column sparsity d .

More formally, on the dimension side, for every $p \in (1, \infty) \setminus \{2\}$, distortion $D > 1$, and (k, D) -RIP- p matrix $A \in \mathbb{R}^{m \times n}$, we show that $m = \Omega(k^p)$, where $\Omega(\cdot)$ hides factors that depend on p and D . Note that, it is not hard to extend an argument of Chandar [Cha10] and obtain a lower bound $m = \Omega(k^{p-1})$.⁴ This additional factor k is exactly what makes our lower bound non-trivial and tight for $p \in (1, \infty) \setminus \{2\}$, and thus enables us to conclude that $p = 2$ is a “singularity”.⁵

As for the column sparsity, we present a simple extension of the argument of Chandar [Cha10] and prove that for every $p \in [1, \infty)$ any (k, D) -RIP- p matrix must have column sparsity $\Omega(k^{p-1})$.

RIP matrices and sparse recovery. We extend the result of Candès, Romberg and Tao [CRT06] to show that, for every $p > 1$, RIP- p matrices allow the stable sparse recovery with the ℓ_p/ℓ_1 guarantee and approximation factors $C_1 = O(k^{-1+1/p})$, $C_2 = O(1)$ in polynomial time. This extension is quite straightforward and seems to be folklore, but, to the best of our knowledge, it is not recorded anywhere.

On the other hand, for every $p \geq 1$, it is almost immediate that *any* matrix A that allows the stable sparse recovery with the ℓ_p/ℓ_1 guarantee—even if it works only for k -sparse signals—*must have the (k, C_2) -RIP- p property*. For the sake of completeness, we have included both the above proofs in Appendix A.

Implications to sparse recovery. Using the above equivalent relationship between the stable sparse recovery problem and the RIP- p matrices, we conclude that the stable sparse recovery with the ℓ_p/ℓ_1 guarantee requires $m = \tilde{\Theta}(k^p)$ measurements for every $p \in [1; \infty) \setminus \{2\}$, and requires $d = \tilde{\Theta}(k^{p-1})$ column sparsity for every $p \in [1, \infty)$. Our results together draw tradeoffs between the following three parameters in stable sparse recovery:

- p , the ℓ_p/ℓ_1 guarantee for the stable sparse recovery,⁶
- m , the number of measurements needed for sketching, and
- d , the running time (per input coordinate) needed for sketching.

It was pointed out by an anonymous referee that for the *noiseless* case—that is, when the noise vector e is always zero—better upper bounds are possible. Using the result of Gilbert *et al.* [GSTV07], one can obtain, for every $p \geq 2$, the noiseless stable sparse recovery procedure with the ℓ_p/ℓ_1 guarantee using only $m = \tilde{O}(k^{2-2/p})$ measurements. Therefore, our results also imply a very large gap, both in terms of m and d , between the *noiseless* and the *noisy* stable sparse recovery problems.

³That is, a (scaled) random $m \times n$ binary matrix with $m = O(\varepsilon^{-2} k^2 \log(n/k))$ rows and sparsity $d = O(\varepsilon^{-1} k \log(n/k))$ satisfies the $(k, 1 + \varepsilon)$ -RIP-2 property. This can be proved using for instance the incoherence argument from [Rau10]: any incoherent matrix satisfies the RIP-2 property with certain parameters.

⁴Also, the same argument gives the lower bound $\Omega(k^p)$ for *binary* RIP- p matrices for every $p \in [1, \infty)$.

⁵A similar singularity is known to exist for linear dimension reduction for arbitrary point sets with respect to ℓ_p norms [LMN05]; alas, tight bounds for that problem are not known.

⁶We note that the ℓ_p/ℓ_1 and the ℓ_q/ℓ_1 guarantees are incomparable. However, it is often more desirable to have larger p in this ℓ_p/ℓ_1 guarantee to ensure a better recovery quality. This is because, if the noise vector $e = 0$, the ℓ_q/ℓ_1 guarantee (with $C_1 = O(k^{-1+1/q})$) can be shown to be stronger than the ℓ_p/ℓ_1 one (with $C_1 = O(k^{-1+1/p})$) whenever $q > p$. However, when there is a noise term, the guarantee $\|x - \hat{x}\|_p \leq O(1) \cdot \|e\|_p$ is incomparable to $\|x - \hat{x}\|_q \leq O(1) \cdot \|e\|_q$ for $p \neq q$.

1.3 Overview of the proofs

Upper bounds. We construct RIP- p matrices as follows. Beginning with a zero matrix A with $m = \tilde{O}(k^p)$ rows and n columns, independently for each column of A , we choose $d = \tilde{O}(k^{p-1})$ out of m entries uniformly at random (without replacement), and assign the value $d^{-1/p}$ to those selected entries. For this construction, we have two very different analyses of its correctness: one works only for $p \geq 2$, and the other works only for $1 < p < 2$.

For $p \geq 2$, the most challenging part is to show that $\|Ax\|_p \leq (1 + \varepsilon)\|x\|_p$ holds with high probability, for all k -sparse vectors x . We reduce this problem to a probabilistic question *similar in spirit* to the following “balls and bins” question. Consider n bins in which we throw n balls uniformly and independently. As a result, we get n numbers X_1, X_2, \dots, X_n , where X_i is the number of balls falling into the i -th bin. We would like to upper bound the tail $\Pr[S \geq 1000 \cdot \mathbb{E}[S]]$ for the random variable $S = \sum_{i=1}^n X_i^{p-1}$. (Here, the constant 1000 can be replaced with any large enough one since we do not care about constant factors in this paper.) The first challenge is that X_i ’s are not independent. To deal with this issue we employ the notion of *negative association* of random variables introduced by Joag-Dev and Proschan [JDP83]. The second problem is that the random variables X_i^{p-1} are heavy tailed: they have tails of the form $\Pr[X_i^{p-1} \geq t] \approx \exp(-t^{\frac{1}{p-1}})$, so the standard technique of bounding the moment-generating function does not work. Instead, we bound the high moments of S directly, which introduces certain technical challenges. Let us remark that sums of i.i.d. heavy-tailed variables were thoroughly studied by Nagaev [Nag69a, Nag69b], but it seems that for the results in these papers the independence of summands is crucial.

One major reason the above approach fails to work for $1 < p < 2$ is that, in this range, even the best possible tail inequality for S is too weak for our purposes. Another challenge in this regime is that, to bound the “lower tail” of $\|Ax\|_p^p$ (that is, to prove that $\|Ax\|_p \geq (1 - \varepsilon)\|x\|_p$ holds for all k -sparse x), the simple argument used for $p \geq 2$ no longer works. Our solution to both problems above is to instead build our RIP matrices based on the following general notion of bipartite expanders.

Definition 1.1. Let $G = (U, V, E)$ with $|U| = n$, $|V| = m$ and $E \subseteq U \times V$ be a bipartite graph such that all vertices from U have the same degree d . We say that G is an (ℓ, d, δ) -expander, if for every $S \subseteq U$ with $|S| \leq \ell$ we have

$$|\{v \in V \mid \exists u \in S \ (u, v) \in E\}| \geq (1 - \delta)d|S|.$$

It is known that random d -regular graphs are good expanders, and we can take the (scaled) adjacency matrix of such an expander and prove that it satisfies the desired RIP- p property for $1 < p < 2$. Our argument can be seen as a subtle interpolation between the argument from [BGI⁺08], which proves that (scaled) adjacency matrices of $(k, d, \Theta(\varepsilon))$ -expanders (with $\tilde{O}(k)$ rows) are $(k, 1 + \varepsilon)$ -RIP-1 and the one using incoherence argument,⁷ which shows that $(2, d, \Theta(\varepsilon/k))$ -expanders give $(k, 1 + \varepsilon)$ -RIP-2 matrices (with $\tilde{O}(k^2)$ rows).

Lower bounds. Our dimension lower bound $m = \Omega(k^p)$ is derived essentially from norm inequalities. The high-level idea can be described in four simple steps. Consider any (k, D) -RIP- p matrix $A \in \mathbb{R}^{n \times m}$, and assume that D is very close to 1 in this high-level description.

In the first three steps, we deduce from the RIP property that (a) the sum of the p -th powers of all entries in A is approximately n , (b) the largest entry in A (i.e., the vector ℓ_∞ -norm of A) is essentially at most $k^{1/p-1}$, and (c) the sum of squares of all entries in A is at least $n(\frac{k}{m})^{2/p-1}$ if

⁷It is known [Rau10] that an incoherent matrix satisfies the RIP-2 property with certain parameters. At the same time, the notion of incoherence can be interpreted as expansion for $\ell = 2$.

$p \in (1, 2)$, or at most $n(\frac{k}{m})^{2/p-1}$ if $p > 2$. In the fourth step, we combine (a) (b) and (c) together by arguing about the relationships between the ℓ_p , ℓ_∞ and ℓ_2 norms of entries of A , and prove the desired lower bound on m .

The sparsity lower bound $d = \Omega(k^{p-1})$ can be obtained via a simple extension of the argument of Chandar [Cha10]. It is possible to extend the techniques of Nelson and Nguyễn [NN13] to obtain a slightly better sparsity lower bound. However, since we were unable to obtain a *tight* bound this way, we decided not to include it.

2 RIP Construction for $p \geq 2$

In this section, we construct $(k, 1 + \varepsilon)$ -RIP- p matrices for $p \geq 2$ by proving the following theorem.

Definition 2.1. *We say that an $m \times n$ matrix A is a random binary matrix with sparsity $d \in [m]$, if A is generated by assigning $d^{-1/p}$ to d random entries per column (selected uniformly at random without replacement), and assigning 0 to the remaining entries.*

Theorem 2.2. *For all $n \in \mathbb{Z}_+$, $k \in [n]$, $\varepsilon \in (0, \frac{1}{2})$ and $p \in [2, \infty)$, there exist $m, d \in \mathbb{Z}_+$ with*

$$m = p^{O(p)} \cdot \frac{k^p}{\varepsilon^2} \cdot \log^{p-1} n \quad \text{and} \quad d = p^{O(p)} \cdot \frac{k^{p-1}}{\varepsilon} \cdot \log^{p-1} n \leq m$$

such that, letting A be a random binary $m \times n$ matrix of sparsity d , with probability at least 98%, A satisfies $(1 - \varepsilon)\|x\|_p^p \leq \|Ax\|_p^p \leq (1 + \varepsilon)\|x\|_p^p$ for all k -sparse vectors $x \in \mathbb{R}^n$.

Our proof is divided into two steps: (1) the “lower-tail step”, that is, with probability at least 0.99 we have $\|Ax\|_p^p \geq (1 - \varepsilon)\|x\|_p^p$ for all k -sparse x , and (2) the “upper-tail step”, that is, with probability at least 0.99, we have $\|Ax\|_p^p \leq (1 + \varepsilon)\|x\|_p^p$.

For every $j \in [n]$, let us denote by $S_j \subseteq [m]$ the set of non-zero rows of the j -th column of A .

2.1 The Lower-Tail Step

The lower-tail step is very simple. It suffices to show that, with high probability, $|S_i \cap S_j|$ is small for every pair of different $i, j \in [n]$, which will then imply that if only k columns of A are considered, every S_i has to be almost disjoint from the union of the S_j of the $k - 1$ remaining columns. This can be summarized by the following claim.

Claim 2.3. *If $d \geq C\varepsilon^{-1}k \log n$ and $m \geq 2dk/\varepsilon$, where C is some large enough constant, then*

$$\Pr \left[\forall 1 \leq i < j \leq n \quad |S_i \cap S_j| \leq \frac{\varepsilon d}{k} \right] \geq 0.99 \quad .$$

Proof. Let us first upper bound the probability that S_i and S_j intersect by more than $\frac{\varepsilon d}{k}$ elements. For notational simplicity suppose that $S_i = \{1, \dots, d\}$, and let the random variable X_k be 1 if S_j contains k , and 0 if not. Under this definition, we have $|S_i \cap S_j| = \sum_{k=1}^d X_k$.

Noticing that the expectation $\mathbb{E}[X_1 + \dots + X_d] = \frac{d}{m} \cdot d = \frac{d^2}{m}$, and $\frac{\varepsilon d}{k} \geq 2 \cdot \frac{d^2}{m}$ is twice as large as the expectation, we apply Chernoff bound for negatively correlated binary random variables [PS97] and obtain

$$\Pr \left[|S_i \cap S_j| > \frac{\varepsilon d}{k} \right] = \Pr \left[X_1 + \dots + X_d > \frac{\varepsilon d}{k} \right] < e^{-\Omega(\varepsilon d/k)} \leq \frac{1}{100n^2} \quad ,$$

where the last inequality is true by our choice of $d \geq C\varepsilon^{-1}k \log n$ for some large enough constant C . Finally, by union bound, we have $\Pr [\exists i, j \in [n] \text{ with } i \neq j, |S_i \cap S_j| > \frac{\varepsilon d}{k}] \leq 0.01$. \square

Now, to prove the lower tail, without loss of generality, let us assume that x is supported on $[k]$, the first k coordinates. For every $j \in [k]$, we denote by $S'_j = S_j \setminus \bigcup_{j' \in [k] \setminus \{j\}} S_{j'}$, the set of non-zero rows in column j that are not shared with the supports of other columns in $[k] \setminus \{j\}$. If the event in Claim 2.3 holds, then for every $j \in [k]$, we have $|S'_j| \geq (1 - \varepsilon)d$. Thus, we can lower bound $\|Ax\|_p$ as

$$\|Ax\|_p^p = \frac{1}{d} \cdot \sum_{i=1}^m \left| \sum_{j \in [k]: i \in S_j} x_j \right|^p \geq \frac{1}{d} \cdot \sum_{i=1}^m \left| \sum_{j \in [k]: i \in S'_j} x_j \right|^p = \frac{1}{d} \cdot \sum_{j \in [k]} |S'_j| \cdot |x_j|^p \geq (1 - \varepsilon) \|x\|_p^p. \quad (2.1)$$

Remark 2.4. *The above claim only works when $m = \Omega(k^2 \log n / \varepsilon^2)$, and therefore we cannot use it in for the case of $1 < p < 2$.*

2.2 The Upper-Tail Step

Below we describe the framework of our proof for the upper-tail step, deferring all technical details to Section 2.2.1.

Suppose again that x is supported on $[k]$. Then, we upper bound $\|Ax\|_p^p$ as

$$\begin{aligned} \|Ax\|_p^p &= \frac{1}{d} \cdot \sum_{i=1}^m \left| \sum_{j \in [k]: i \in S_j} x_j \right|^p \leq \frac{1}{d} \cdot \sum_{i=1}^m |\{j' \in [k] \mid i \in S_{j'}\}|^{p-1} \cdot \sum_{j \in [k]: i \in S_j} |x_j|^p \\ &= \frac{1}{d} \cdot \sum_{j=1}^k |x_j|^p \cdot \sum_{i \in S_j} |\{j' \in [k] \mid i \in S_{j'}\}|^{p-1}, \end{aligned} \quad (2.2)$$

where the first inequality follows from the fact that $(a_1 + \dots + a_N)^p \leq N^{p-1}(a_1^p + \dots + a_N^p)$ for any sequence of N non-negative reals a_1, \dots, a_N . Note that the quantity $|\{j' \in [k] \mid i \in S_{j'}\}| \in [k]$ captures the number of non-zeros of A in the i -th row and the first k columns. From now on, in order to prove the desired upper tail, it suffices to show that, with high probability

$$\forall j \in [k], \quad \sum_{i \in S_j} |\{j' \in [k] \mid i \in S_{j'}\}|^{p-1} \leq (1 + \varepsilon)d. \quad (2.3)$$

To prove this, let us fix some $j^* \in [k]$ and upper bound the probability that (2.3) holds for $j = j^*$, and then take a union bound over the choices of j^* . Without loss of generality, assume that $S_{j^*} = \{1, 2, \dots, d\}$, consisting of the first d rows. For every $i \in S_{j^*}$, define a random variable $X_i \stackrel{\text{def}}{=} |\{j' \in [k] \mid i \in S_{j'}\}| - 1$. It is easy to see that X_i is distributed as $\text{Bin}(k - 1, d/m)$, the binomial distribution that is the sum of $k - 1$ i.i.d. random 0/1 variables, each being 1 with probability d/m . For notational simplicity, let us define $\delta \stackrel{\text{def}}{=} dk/m$. We will later choose $\delta < \varepsilon$ to be very small. Our goal in (2.3) can now be reformulated as follows: upper bound the probability

$$\Pr \left[\sum_{i=1}^d ((X_i + 1)^{p-1} - 1) > \varepsilon d \right].$$

We begin with a lemma showing an upper bound on the moments of each $Y_i \stackrel{\text{def}}{=} (X_i + 1)^{p-1} - 1$.

Lemma 2.5. *There exists a constant $C \geq 1$ such that, if X is drawn from the binomial distribution $\text{Bin}(k - 1, \delta/k)$ for some $\delta < 1/(2e^2)$, and $p \geq 2$, then for any real $\ell \geq 1$,*

$$\mathbb{E}[(X + 1)^{p-1} - 1]^\ell \leq C \cdot \delta(\ell(p - 1) + 1)^{\ell(p-1)+1}.$$

Next, we note that although the random variables X_i 's are dependent, they can be verified to be *negatively associated*, a notion introduced by Joag-Dev and Proschan [JDP83]. This theory allows us to conclude the following bound on the moments.

Lemma 2.6. Let $\tilde{X}_1, \dots, \tilde{X}_d$ be d random variables, each drawn independently from $\text{Bin}(k-1, \delta/k)$. Then, for every integer $t \geq 1$ we have

$$\mathbb{E} \left[\left(\sum_{i=1}^d ((X_i + 1)^{p-1} - 1) \right)^t \right] \leq \mathbb{E} \left[\left(\sum_{i=1}^d ((\tilde{X}_i + 1)^{p-1} - 1) \right)^t \right] .$$

Now, using the moments of random variables $Y_i = (X_i + 1)^{p-1} - 1$ from Lemma 2.5, as well as Lemma 2.6, we can compute the tail bound of the sum $\sum_{i=1}^d Y_i$. Our proof of the following Lemma uses the result of Latała [Lat97].

Lemma 2.7. There exists constants $C \geq 1$ such that, whenever $\delta \leq \varepsilon/p^{Cp}$ and $d \geq p^{Cp}/\varepsilon$, we have

$$\Pr \left[\sum_{i=1}^d ((X_i + 1)^{p-1} - 1) > \varepsilon d \right] \leq e^{-\Omega\left(\frac{(\varepsilon d)^{1/(p-1)}}{p}\right)} .$$

Finally, we are ready to prove Theorem 2.2.

Proof of Theorem 2.2. We can choose $d = \Theta(p)^{p-1} \cdot \frac{k^{p-1}}{\varepsilon} \cdot \log^{p-1} n$ so that $e^{-\Omega\left(\frac{(\varepsilon d)^{1/(p-1)}}{p}\right)} < \frac{1}{100} \frac{1}{k \binom{n}{k}}$.

Since our choice of $m = \frac{dkp^{\Theta(p)}}{\varepsilon}$ ensures that $\delta = dk/m \leq \varepsilon/p^{Cp}$, and our choice of d ensures $d \geq p^{Cp}/\varepsilon$, we can apply Lemma 2.7 and conclude that with probability at least $1 - \frac{1}{100} \frac{1}{k \binom{n}{k}}$ one has

$$\sum_{i \in S_{j^*}} |\{j' \in [k] \mid i \in S_{j'}\}|^{p-1} = \sum_{i=1}^d (X_i + 1)^{p-1} \leq (1 + \varepsilon)d .$$

Therefore, by applying the union bound over all $j^* \in [k]$, we conclude that with probability at least $1 - \frac{1}{100} \frac{1}{\binom{n}{k}}$, the desired inequality (2.3) is satisfied for all $j \in [k]$.

Recall that, owing to (2.2), the inequality (2.3) implies that $\|Ax\|_p^p \leq (1 + \varepsilon)\|x\|_p^p$ for every $x \in \mathbb{R}^n$ that is supported on the *first* k coordinates. By another union bound over the choices of all possible $\binom{n}{k}$ subsets of $[n]$, we conclude that with probability at least 0.99, we have $\|Ax\|_p^p \leq (1 + \varepsilon)\|x\|_p^p$ for all k -sparse vectors x .

On the other hand, since our choice of d and m satisfies the assumptions $d \geq \Omega(k \log n/\varepsilon)$ and $m \geq 2dk/\varepsilon$ in Claim 2.3, the lower tail $\|Ax\|_p^p \geq (1 - \varepsilon)\|x\|_p^p$ also holds with probability at least 0.99. Overall we conclude that with probability at least 0.98, we have $\|Ax\|_p^p \in (1 \pm \varepsilon)\|x\|_p^p$ for every k -sparse vector $x \in \mathbb{R}^n$. \square

2.2.1 Missing Proofs

Lemma 2.5. There exists some constant $C \geq 1$ such that, if X is drawn from the binomial distribution $\text{Bin}(k-1, \delta/k)$ for some $\delta < 1/(2e^2)$, and $p \geq 2$, then for any real $\ell \geq 1$,

$$\mathbb{E}[(X + 1)^{p-1} - 1]^\ell \leq C \cdot \delta(\ell(p-1) + 1)^{\ell(p-1)+1} .$$

Proof. We first expand the expectation using the definition of $\text{Bin}(k-1, \delta/k)$.

$$\begin{aligned} \mathbb{E}[(X + 1)^{p-1} - 1]^\ell &= \sum_{i=0}^{k-1} ((i + 1)^{p-1} - 1)^\ell \binom{k-1}{i} \left(1 - \frac{\delta}{k}\right)^{k-1-i} \left(\frac{\delta}{k}\right)^i \\ &\leq \sum_{i=1}^{k-1} ((i + 1)^{p-1} - 1)^\ell \left(\frac{e(k-1)}{i}\right)^i \left(\frac{\delta}{k}\right)^i \\ &\leq \sum_{i=1}^{\infty} (i + 1)^{\ell(p-1)} \left(\frac{e\delta}{i}\right)^i . \end{aligned}$$

Let us denote by $a_i \stackrel{\text{def}}{=} (i+1)^{\ell(p-1)} \left(\frac{e\delta}{i}\right)^i$ the i -th term of the above infinite sum. We have

$$\frac{a_i}{a_{i-1}} = \left(1 + \frac{1}{i}\right)^{\ell(p-1)} \frac{e\delta}{i} \left(1 - \frac{1}{i}\right)^{i-1} \leq \delta \cdot e^{\ell(p-1)/i+1}.$$

Since $\delta < 1/(2e^2)$, we have $a_i/a_{i-1} < 1/2$ for every $i \geq \max\{\ell(p-1), 2\}$. Therefore, the largest $\max_{i \geq 1} a_i$ is obtained when $i = i^* < \max\{\ell(p-1), 2\}$, which implies $1 \leq i^* \leq \ell(p-1)$ because $\ell(p-1) \geq 1$ (here we crucially use that $p \geq 2$ and $\ell \geq 1$). Therefore,

$$\max_{i \geq 1} a_i \leq \left(\frac{e\delta}{i^*}\right)^{i^*} \cdot (\ell(p-1) + 1)^{\ell(p-1)} \leq e\delta \cdot (\ell(p-1) + 1)^{\ell(p-1)},$$

and the second inequality is because $e\delta < 1$. Overall,

$$\mathbb{E}[(X+1)^{p-1} - 1]^\ell \leq \sum_{i=1}^{\infty} a_i \leq \left(\ell(p-1) + \sum_{j=1}^{\infty} 2^{-j}\right) \cdot \max_i a_i \leq O(\delta) \cdot (\ell(p-1) + 1)^{\ell(p-1)+1}. \quad \square$$

Lemma 2.6. *Letting $\tilde{X}_1, \dots, \tilde{X}_d$ be d random variables, each drawn independently from $\text{Bin}(k-1, \delta/k)$. Then, for every integer $t \geq 1$ we have*

$$\mathbb{E} \left[\left(\sum_{i=1}^d ((X_i + 1)^{p-1} - 1) \right)^t \right] \leq \mathbb{E} \left[\left(\sum_{i=1}^d ((\tilde{X}_i + 1)^{p-1} - 1) \right)^t \right].$$

Proof. This lemma follows from the theory of *negatively associated* random variables [JDP83] (see also [Efr65]).

For every $i \in [m]$ and $j \in [k]$, let the random variable $Z_{ij} = 1$ if $i \in S_j$ and 0 otherwise. The random variables across columns are independent: that is, $\{Z_{1,j}, \dots, Z_{m,j}\}$ and $\{Z_{1,j'}, \dots, Z_{m,j'}\}$ are independent if $j \neq j'$.

However, within a single column, Z_{1j}, \dots, Z_{mj} are not independent. In fact, this distribution can be seen as the uniform distribution over $\{0, 1\}^m$ conditioned on $Z_{1j} + \dots + Z_{mj} = d$. Thus, applying [JDP83, Theorem 2.8], we have that $Z_{1j}, Z_{2j}, \dots, Z_{mj}$ are negatively associated (this uses the fact that the Bernoulli distribution is a *Pólya frequency function of order two*). Now, combining the k independent columns, we get that the variables $\{Z_{ij}\}_{i \in [m], j \in [k]}$ are negatively associated altogether.

Next, we want to show that the variables $\{X_1, \dots, X_d\}$ are also negatively associated. By definition, we have $X_i = \sum_{j \neq j^*} Z_{i,j}$. Therefore, X_1, \dots, X_d is a sequence of random variables, each being a partial sum of $\{Z_{ij}\}_{i \in [m], j \in [k]}$, and different X_i 's cover disjoint subsets of $\{Z_{ij}\}_{i \in [m], j \in [k]}$. Applying [JDP83, Property P_6], we have that the variables $\{X_1, \dots, X_d\}$ are negatively associated.

Finally, since the function $f(x) = (x+1)^{p-1} - 1$ is non-decreasing for $p \geq 1$, we apply [JDP83, Property P_6] and conclude that the variables $\{(X_i + 1)^{p-1} - 1\}_{i \in [d]}$ are also negatively associated. Letting $Y_i \stackrel{\text{def}}{=} (X_i + 1)^{p-1} - 1$, then [JDP83, Property P_2] gives that $\mathbb{E}[\prod_{i \in S} Y_i^{r_i}] \leq \prod_{i \in S} \mathbb{E}[Y_i^{r_i}]$ for any subset $S \subseteq [d]$ and any sequence of powers $r_1, \dots, r_d \in \mathbb{Z}_{\geq 0}$.

As a result, we conclude that, letting $\tilde{Y}_i \stackrel{\text{def}}{=} (\tilde{X}_i + 1)^{p-1} - 1$,

$$\begin{aligned} \mathbb{E} \left[\left(\sum_{i=1}^d Y_i \right)^t \right] &= \mathbb{E} \left[\sum_{\substack{r_1, \dots, r_d \in \{0, 1, \dots, d\} \\ r_1 + \dots + r_d = t}} \binom{t}{r_1, \dots, r_d} Y_1^{r_1} Y_2^{r_2} \dots Y_d^{r_d} \right] \\ &\leq \sum_{r_1 + \dots + r_d = t} \binom{t}{r_1, \dots, r_d} \mathbb{E}[Y_1^{r_1}] \mathbb{E}[Y_2^{r_2}] \dots \mathbb{E}[Y_d^{r_d}] \\ &= \sum_{r_1 + \dots + r_d = t} \binom{t}{r_1, \dots, r_d} \mathbb{E}[\tilde{Y}_1^{r_1}] \mathbb{E}[\tilde{Y}_2^{r_2}] \dots \mathbb{E}[\tilde{Y}_d^{r_d}] = \mathbb{E} \left[\left(\sum_{i=1}^d \tilde{Y}_i \right)^t \right]. \quad \square \end{aligned}$$

To prove Lemma 2.7, we need the following theorem of Latała on the moments of sums of i.i.d. non-negative random variables.

Theorem 2.8 ([Lat97], Theorem 1). *Let Y_1, \dots, Y_d be a sequence of independent non-negative random variables from distribution \mathcal{D} and $t \geq 1$. Then*

$$\mathbb{E}[(Y_1 + \dots + Y_d)^t]^{1/t} \leq e \cdot \inf \left\{ u > 0 : \mathbb{E}_{Y \sim \mathcal{D}} \left[\left(1 + \frac{Y}{u} \right)^t \right] \leq e^{t/d} \right\}.$$

Lemma 2.7. *There exists a constant $C \geq 1$ such that, whenever $\delta \leq \varepsilon/p^{Cp}$ and $d \geq p^{Cp}/\varepsilon$, we have*

$$\Pr \left[\sum_{i=1}^d ((X_i + 1)^{p-1} - 1) > \varepsilon d \right] \leq e^{-\Omega(\frac{(\varepsilon d)^{1/(p-1)}}{p})}.$$

Proof. Denote by Y the random variable whose value $Y = (X + 1)^{p-1} - 1$, where X is drawn from the binomial distribution $\text{Bin}(k-1, \delta/k)$, and \mathcal{D} the distribution for Y . We wish to apply Theorem 2.8 for the case of d independent samples from \mathcal{D} , and let us compute the value of u from the statement of Theorem 2.8 as follows. For every integer $t \geq 1$,

$$\mathbb{E}_{Y \sim \mathcal{D}} \left[\left(1 + \frac{Y}{u} \right)^t \right] = 1 + \sum_{\ell=1}^t \binom{t}{\ell} \frac{\mathbb{E}[Y^\ell]}{u^\ell} \leq 1 + \sum_{\ell=1}^t \left(\frac{et}{\ell} \right)^\ell \frac{\mathbb{E}[Y^\ell]}{u^\ell}.$$

This sum, owing to Lemma 2.5, can be upper bounded as

$$\begin{aligned} \sum_{\ell=1}^t \left(\frac{et}{u\ell} \right)^\ell \mathbb{E}[Y^\ell] &\leq O(\delta) \cdot \sum_{\ell=1}^t \left(\frac{et}{u\ell} \right)^\ell \cdot (\ell(p-1) + 1)^{\ell(p-1)+1} \leq O(\delta) \cdot \sum_{\ell=1}^t \left(\frac{et}{u\ell} \right)^\ell \cdot (\ell p)^{\ell(p-1)+1} \\ &\leq O(\delta p) \cdot \sum_{\ell=1}^t \ell \left(\frac{et}{u\ell} \right)^\ell \cdot (\ell p)^{\ell(p-1)} = O(\delta p) \cdot \sum_{\ell=1}^t \ell \left(\frac{ep^{p-1} \cdot t\ell^{p-2}}{u} \right)^\ell \\ &= O\left(\frac{\delta p^p t}{u} \right) \cdot \sum_{\ell=1}^t \ell^{p-1} \left(\frac{ep^{p-1} \cdot t\ell^{p-2}}{u} \right)^{\ell-1} \leq O\left(\frac{\delta p^p t}{u} \right) \cdot \sum_{\ell=1}^t \ell^{p-1} \left(\frac{ep^{p-1} \cdot t^{p-1}}{u} \right)^{\ell-1}. \end{aligned}$$

Above, the last inequality has used the fact that $p \geq 2$. Now, by choosing $u \geq 2ep^{p-1} \cdot t^{p-1}$ we have that

$$\sum_{\ell=1}^t \left(\frac{et}{u\ell} \right)^\ell \mathbb{E}[Y^\ell] \leq O\left(\frac{\delta p^p t}{u} \right) \cdot \sum_{\ell=1}^t \ell^{p-1} \frac{1}{2^{\ell-1}}. \quad (2.4)$$

Since

$$\sum_{\ell=1}^{\infty} \ell^{p-1} \frac{1}{2^{\ell-1}} \leq 2 \cdot \int_0^{\infty} x^{p-1} \cdot 2^{-x} dx \leq 2^{O(p)} \cdot \Gamma(p) \leq p^{O(p)},$$

we conclude that the right hand side of (2.4) is upper bounded by $O(\frac{\delta p^{O(p)} t}{u})$. In sum, we conclude that when $u = 2ep^{p-1} \cdot t^{p-1}$ and $t^{p-1} \geq \delta p^{\Omega(p)} d$, we have

$$\mathbb{E}_{Y \sim \mathcal{D}} \left[\left(1 + \frac{Y}{u} \right)^t \right] \leq 1 + O\left(\frac{\delta p^{O(p)} t}{u} \right) \leq 1 + \frac{t}{d} < e^{t/d}.$$

Invoking Theorem 2.8 for this choice of $u = 2ep^{p-1} \cdot t^{p-1}$ and for any integer $t \geq 1$ satisfying $t^{p-1} \geq \delta p^{\Omega(p)} d$, we have

$$\mathbb{E} \left[\left(\sum_{i=1}^d ((\tilde{X}_i + 1)^{p-1} - 1) \right)^t \right]^{1/t} \leq 2e^2 p^{p-1} \cdot t^{p-1},$$

where each \tilde{X}_i is an i.i.d. sample from $\text{Bin}(k-1, \delta/k)$. Invoking Lemma 2.6, we obtain the same moment bound on X_1, \dots, X_d .

$$\mathbb{E} \left[\left(\sum_{i=1}^d ((X_i + 1)^{p-1} - 1) \right)^t \right]^{1/t} \leq 2e^2 p^{p-1} \cdot t^{p-1}.$$

Using Markov's inequality, we have for any integer $t \geq 1$ satisfying $t^{p-1} \geq \delta p^{\Omega(p)} d$,

$$\Pr \left[\sum_{i=1}^d ((X_i + 1)^{p-1} - 1) > \varepsilon d \right] \leq \frac{1}{(\varepsilon d)^t} \mathbb{E} \left[\left(\sum_{i=1}^d ((X_i + 1)^{p-1} - 1) \right)^t \right] \leq \left(\frac{2e^2 p^{p-1} \cdot t^{p-1}}{\varepsilon d} \right)^t.$$

By the assumption $d \geq p^{Cp}/\varepsilon$, so let us choose t to be the largest positive integer such that $\frac{2e^2 p^{p-1} \cdot t^{p-1}}{\varepsilon d} \leq \frac{1}{2}$. That is, $t = \Theta\left(\frac{(\varepsilon d)^{1/(p-1)}}{p}\right)$. Since $\delta < \varepsilon/p^{Cp}$, we have $t^{p-1} \geq \delta p^{\Omega(p)} d$. Thus,

$$\Pr \left[\sum_{i=1}^d ((X_i + 1)^{p-1} - 1) > \varepsilon d \right] \leq 2^{-\Omega(t)} \leq e^{-\Omega\left(\frac{(\varepsilon d)^{1/(p-1)}}{p}\right)}. \quad \square$$

3 RIP Construction for $1 < p < 2$

In this section, we construct $(k, 1 + \varepsilon)$ -RIP- p matrices for $1 < p < 2$ by proving the following theorem.

We assume that $1 + \tau \leq p \leq 2 - \tau$ for some $\tau > 0$, and whenever we write $O_\tau(\cdot)$, we assume that some factor that depends on τ is hidden. (For instance, factors of $p/(1-p)$ may be hidden.)

Theorem 3.1. *For every $n \in \mathbb{Z}_+$, $k \in [n]$, $0 < \varepsilon < 1/2$ and $1 + \tau \leq p \leq 2 - \tau$, there exist $m, d \in \mathbb{Z}_+$ with*

$$m = O_\tau \left(k^p \frac{\log n}{\varepsilon^2} + k^{4-2/p-p} \frac{\log n}{\varepsilon^{2/(p-1)}} \right) \quad \text{and} \quad d = O_\tau \left(\frac{k^{p-1} \cdot \log n}{\varepsilon} + \frac{k^{(p-1)/p} \cdot \log n}{\varepsilon^{1/(p-1)}} \right)$$

such that, letting A be a random binary $m \times n$ matrix of sparsity d , with probability at least 98%, A satisfies $(1 - \varepsilon)\|x\|_p^p \leq \|Ax\|_p^p \leq (1 + \varepsilon)\|x\|_p^p$ for all k -sparse vectors $x \in \mathbb{R}^n$.

Note that, when $k \geq \varepsilon^{-\frac{p(2-p)}{(p-1)^3}}$, the above bounds on m and k can be simplified as

$$m = O_\tau \left(\frac{k^p \cdot \log n}{\varepsilon^2} \right) \quad \text{and} \quad d = O_\tau \left(\frac{k^{p-1} \cdot \log n}{\varepsilon} \right).$$

Our proof of the above theorem is based on the existence of (ℓ, d, δ) bipartite expanders (recall the definition of such expanders from Definition 1.1):

Lemma 3.2. *[BMRV02, Lemma 3.10] For every $\delta \in (0, \frac{1}{2})$, and $\ell \in [n]$, there exist (ℓ, d, δ) -expanders with $d = O\left(\frac{\log n}{\delta}\right)$ and $m = O(d\ell/\delta) = O\left(\frac{\ell \log n}{\delta^2}\right)$.*

In fact, the proof of Lemma 3.2 implies a simple probabilistic construction of such expanders: with probability at least 98%, a random binary matrix A of sparsity d is the adjacency matrix of a $(2\ell, d, \delta)$ -expander scaled by $d^{-1/p}$, for $\delta = \Theta\left(\frac{\log n}{d}\right)$ and $\ell = \Theta\left(\frac{\delta m}{d}\right)$.

Therefore, we will assume that A is the (scaled) adjacency matrix of a $(2\ell, d, \delta)$ -expander, for parameters of ℓ and δ that we will specify in the end of this section.⁸

⁸In fact, we will choose $\ell = \Theta_\tau(k^{2-p})$. Therefore, our construction confirms our description in the introduction: it interpolates between the expander construction of RIP-1 matrices from [BGI⁺08] that uses $\ell = k$, and the construction of RIP-2 matrices using incoherence argument that essentially corresponds to $\ell = 2$.

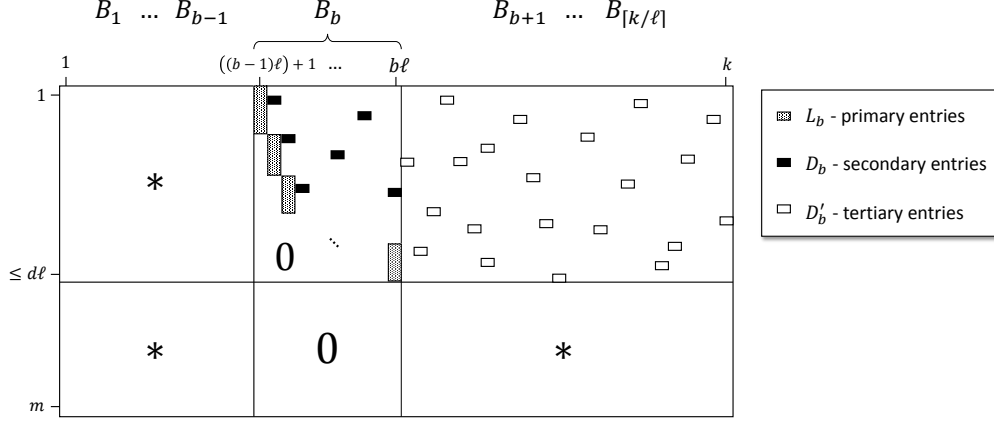


Figure 1: Illustrating the definitions of $B_b \subseteq [k]$, L_b , D_b and D'_b .

3.1 High-Level Proof Idea

The goal is to show that $|\|Ax\|_p^p - 1| \leq \varepsilon$ for every k -sparse vector x that satisfies $\|x\|_p = 1$. Without loss of generality, let us assume that x is supported on $[k]$, the first k coordinates among $[n]$, and $|x_1| \geq |x_2| \geq \dots \geq |x_k|$.

We partition the k columns into $\lceil k/\ell \rceil$ blocks each of size ℓ , and denote them by $B_1 = \{1, 2, \dots, \ell\}$, $B_2 = \{\ell + 1, \ell + 2, \dots, 2\ell\}$, and so on. With this definition, we can expand $\|Ax\|_p^p$ as follows:

$$\begin{aligned} |\|Ax\|_p^p - 1| &= \left| \sum_{i=1}^m \left| \sum_{j=1}^k A_{ij} x_j \right|^p - \|x\|_p^p \right| = \left| \sum_{i=1}^m \left| \sum_{j=1}^k A_{ij} x_j \right|^p - \sum_{i=1}^m \sum_{j=1}^k |A_{ij} x_j|^p \right| \\ &\leq O(1) \cdot \sum_{i=1}^m \sum_{j=1}^k \left(|A_{ij} x_j| \cdot \left| \sum_{j'=j+1}^k A_{ij'} x_{j'} \right|^{p-1} \right) = O(1) \cdot \sum_{b=1}^{\lceil k/\ell \rceil} \sum_{i=1}^m \sum_{j \in B_b} \left(|A_{ij} x_j| \cdot \left| \sum_{j'=j+1}^k A_{ij'} x_{j'} \right|^{p-1} \right), \end{aligned} \quad (3.1)$$

where the inequality follows from Claim 3.3, a tight bound on the difference between ‘the p -th power of the sum’ and ‘the sum of the p -th powers’.

To upper bound the right-hand side of (3.1), we fix a block $B_b = \{(b-1)\ell + 1, \dots, b\ell\}$ and consider three groups of non-zero entries of A : ‘primary’, ‘secondary’ and ‘tertiary’ entries.

Let us first define primary and secondary entries: together they form a partition of non-zero entries in the columns of the block B_b . We define *primary* entries $L_b \subseteq [m] \times B_b$ using the following procedure. For every row of A that has non-zero entries in the columns of B_b , we pick the non-zero entry with the smallest column index and add it to the set of primary entries L_b . We define *secondary* entries $D_b \subseteq [m] \times B_b$ to be the remaining non-zero entries in the columns of B_b . Finally, we define *tertiary* entries $D'_b \subseteq [m] \times (B_{b+1} \cup \dots \cup B_{\lceil k/\ell \rceil})$ as the set of non-zero entries that lie in the same row as some primary entry from L_b and in some block $B_{b'}$ for $b' > b$ (see Figure 1, where we permute rows of A for the sake of illustration).

Next, let us sketch how we upper bound the right-hand side of (3.1). First, along the way we use crucially the simple estimate $|x_j| \leq j^{-1/p}$ for every $j \in [k]$. Second, we upper bound the following partial sum of (3.1) for each b separately:

$$\sum_{i=1}^m \sum_{j \in B_b} \left(|A_{ij} x_j| \cdot \left| \sum_{j'=j+1}^k A_{ij'} x_{j'} \right|^{p-1} \right).$$

We further decompose this sum with respect to (i, j) that are primary (i.e., in L_b) or secondary (i.e., in D_b), and notice that the pairs (i, j') are either secondary or tertiary (i.e., in $D_b \cup D'_b$). The crucial observation in our proof is that the entries in $D_b \cup D'_b$ are very sparse and spread across the columns due to the expansion property of A . Another observation is that for L_b , we have at most d entries per column, so we can control the magnitudes of $|x_j| \leq j^{-1/p}$ for $(i, j) \in L_b$ fairly well. Overall, the proof of upper bounding the right hand side of (3.1) boils down to the careful exploitation of these observations and several applications of Hölder's inequality. The details are somewhat lengthy: in particular, we have to treat the case $b = 1$ separately, and carefully choose all the parameters. The rest of this section contains the full analysis of this high-level proof idea.

3.2 Preliminaries

Claim 3.3. *There exists an absolute positive constant $C > 0$ such that⁹ for every $a, b \in \mathbb{R}$ and $1 \leq p \leq 2$ one has*

$$||a + b|^p - |a|^p - |b|^p| \leq C|a||b|^{p-1} . \quad (3.2)$$

Therefore, by induction, we obtain that for every $a_1, a_2, \dots, a_n \in \mathbb{R}$ and $1 \leq p \leq 2$, it satisfies that

$$\left| \sum_{i=1}^n a_i \right|^p \in \sum_{i=1}^n |a_i|^p \pm C \cdot \sum_{i=1}^{n-1} |a_i| \cdot \left| \sum_{j=i+1}^n a_j \right|^{p-1} .$$

Proof. If $a = 0$ or $b = 0$, then (3.2) is true for any $C > 0$. Otherwise, by homogeneity we can assume that $a = 1$. We prove (3.2) separately for $b < 0$ and $b > 0$.

The Case $b < 0$. First, consider the case $b < 0$. Our goal is to prove that

$$\frac{1 + |b|^p - |1 + b|^p}{|b|^{p-1}} \quad (3.3)$$

is bounded from above by a constant for $b < 0$ and $1 \leq p \leq 2$ (obviously, (3.3) is non-negative). Since (3.3) is continuous for these values of b and p , it is sufficient to prove that (3.3) is bounded in each of the following two cases:

- $b < b_1$ and $1 \leq p \leq 2$;
- $b_2 < b < 0$ and $1 \leq p \leq 2$,

where $b_1 < b_2 < 0$ are arbitrary constants (one should think of b_1 having a large absolute value and b_2 being close to zero).

First, let us consider the case $b < b_1$. Denoting by $x = -b > 0$, we need to prove that for sufficiently large x and $1 \leq p \leq 2$,

$$\frac{1 + x^p - (x - 1)^p}{x^{p-1}}$$

is bounded. We have

$$\frac{1 + x^p - (x - 1)^p}{x^{p-1}} = \frac{1 + x^p - x^p \cdot (1 - 1/x)^p}{x^{p-1}} \leq \frac{1 + x^p - x^p \cdot (1 - p/x)}{x^{p-1}} \leq p + 1 ,$$

where the first inequality is due to the generalized Bernoulli's inequality, and the second inequality holds, if x is sufficiently large.

⁹In fact, choosing $C = 3$ should suffice for this claim, but that will make the proof significantly longer.

Now, let us consider the case $b_2 < b < 0$. Defining $\varepsilon = -b > 0$, we need to prove that for sufficiently small $\varepsilon > 0$ and $1 \leq p \leq 2$,

$$\frac{1 + \varepsilon^p - (1 - \varepsilon)^p}{\varepsilon^{p-1}}$$

is bounded. We have

$$\frac{1 + \varepsilon^p - (1 - \varepsilon)^p}{\varepsilon^{p-1}} \leq \frac{1 + \varepsilon^p - (1 - p\varepsilon)}{\varepsilon^{p-1}} \leq \frac{(p+1)\varepsilon}{\varepsilon^{p-1}} \leq p+1,$$

where the first inequality is due to the generalized Bernoulli's inequality, the second and third inequalities follow from the fact that $1 \leq p \leq 2$.

The Case $b > 0$. Let us now handle the case $b > 0$. It is sufficient to check that for every $b \geq 0$ and $1 \leq p \leq 2$ we have

$$(1+b)^p \leq 1 + b^p + pb^{p-1}.$$

This inequality is trivially true when $b = 0$, and therefore, it is enough to check that for every $b > 0$ and $1 \leq p \leq 2$,

$$\frac{\partial}{\partial b}((1+b)^p - 1 - b^p - pb^{p-1}) = p(1+b)^{p-1} - pb^{p-1} - p(p-1)b^{p-2} \leq 0$$

or equivalently

$$\left(1 + \frac{1}{b}\right)^{p-1} \leq 1 + \frac{p-1}{b}.$$

But the latter follows from the generalized Bernoulli's inequality. \square

Lemma 3.4. *For every $a, c \in \mathbb{R}_{\geq 0}^N$ and $1 \leq p \leq 2$, we have*

$$\sum_{i=1}^N c_i a_i^{p-1} \leq \|c\|_{1/(2-p)} \cdot \|a\|_1^{p-1}.$$

In particular, if $c_1 = \dots = c_N = 1$, we have

$$\sum_{i=1}^N a_i^{p-1} \leq N^{2-p} \cdot \left(\sum_{i=1}^N a_i\right)^{p-1}.$$

Proof. This is just an application of Hölder's inequality for norms $1/(2-p)$ and $1/(p-1)$. \square

3.3 From Expansion Property to the Primary-Secondary-Tertiary Decomposition

Using the notation from Section 3.1, let us translate the expansion property into a cardinality upper bound on the sets of secondary and tertiary entries.

Lemma 3.5. *For every integer $1 \leq b \leq \lceil k/\ell \rceil$, we have for every integer $t \geq 1$,*

$$\left| \{(i, j) \in D_b \cup D'_b \mid j \leq (b-1)\ell + t\} \right| \leq \begin{cases} 0, & t = 1, \\ 3\delta dt, & t > 1. \end{cases} \quad (3.4)$$

In addition, we have $|D_b| \leq \delta d\ell$.

Proof. First of all, $|D_b| \leq \delta d\ell$ is an immediate corollary of the expansion property. Recall that A is the (scaled) adjacency matrix of a $(2\ell, d, \delta)$ -expander and therefore $|D_b| = d|B_b| - |\bigcup_{j \in B_b} S_j| \leq d|B_b| - (1 - \delta)d|B_b| = \delta d|B_b| \leq \delta d\ell$.

(3.4) for $t = 1$ is obvious, because in the column of $(b - 1)\ell + 1$, there are only primary entries but not secondary or tertiary ones (see Figure 1).

For any integer t between 2 and ℓ , we observe that the left hand side of (3.4) consists only of secondary entries in D_b , and moreover,

$$\left| \{(i, j) \in D_b \mid j \leq (b - 1)\ell + t\} \right| = dt - \left| \bigcup_{j=(b-1)\ell+1}^{j=(b-1)\ell+t} S_j \right| \leq dt - (1 - \delta)dt = \delta dt \leq 3\delta dt .$$

For any $t > \ell$, we argue as follows. Since the expander property of A ensures that the union of any 2ℓ distinct S_j 's have at least $(1 - \delta)2d\ell$ distinct elements, we conclude that for every $b^* > b$:

$$\left| \{(i, j) \in D'_b \mid j \in B_{b^*}\} \right| \leq 2d\ell - \left| \bigcup_{j \in B_b \cup B_{b^*}} S_j \right| \leq 2\delta d\ell .$$

Therefore, for any integer $t > \ell$, suppose that $(b - 1)\ell + t \in B_{b'}$ for $b' = b + \lfloor \frac{t-1}{\ell} \rfloor > b$, then we have

$$\left| \{(i, j) \in D_b \cup D'_b \mid j \leq (b - 1)\ell + t\} \right| \leq |D_b| + (b' - b) \cdot 2\delta d\ell \leq \delta d\ell + \frac{t-1}{\ell} \cdot 2\delta d\ell \leq 3\delta dt .$$

This finishes all the cases of Lemma 3.5. \square

The expansion property implies the following useful inequality that will be used extensively in the proof.

Lemma 3.6. *For every integer $1 \leq b \leq \lceil k/\ell \rceil$, we have*

$$\sum_{(i,j) \in D_b \cup D'_b} A_{ij}|x_j| \leq 3\delta(dk)^{1-1/p} =: S. \quad (3.5)$$

We denote by $S = 3\delta(dk)^{1-1/p}$ the right-hand side of (3.5).

Proof. ¹⁰ Since each non-zero entry of A equals to $d^{-1/p}$, the left hand side of the desired inequality is

$$d^{-1/p} \cdot \sum_{(i,j) \in D_b \cup D'_b} |x_j| = d^{-1/p} \cdot \sum_{j \geq (b-1)\ell+1} |x_j| \cdot \left| \{i \mid (i, j) \in D_b \cup D'_b\} \right| .$$

Let us denote by $a_j = |\{i \mid (i, j) \in D_b \cup D'_b\}|$, the number of distinct nonzero elements in the j -th column of A that share rows with the primary entries L_b of the block b . Then, the above sum equals to

$$d^{-1/p} \cdot \sum_{j \geq (b-1)\ell+1} |x_j| \cdot a_j = d^{-1/p} \cdot \sum_{t \geq 1} |x_{(b-1)\ell+t}| \cdot a_{(b-1)\ell+t} .$$

We now observe that, $a_{(b-1)\ell+1} + \dots + a_{(b-1)\ell+t} \leq 3\delta dt$ for every $t \geq 1$ according to Lemma 3.5, while at the same time, $|x_{(b-1)\ell+t}|$ is assumed to be non-increasing as t increases. Therefore, it one can see that the right hand side of the above sum is maximized when

$$a_{(b-1)\ell+1} = \dots = a_{(b-1)\ell+t} = \dots = 3\delta d ,$$

¹⁰This is a simple modification of [BGI⁺08, Lemma 9]. However, that lemma does not directly apply for our scenario, because it assumes A expanding any subsets of size at most k .

and therefore, we conclude that

$$\sum_{(i,j) \in D_b \cup D'_b} A_{ij} |x_j| \leq d^{-1/p} \cdot 3\delta d \cdot \|x\|_1 \leq 3\delta (dk)^{1-1/p},$$

where the last inequality follows from the relation between ℓ_1 and ℓ_p norms, that is $\|x\|_1 \leq k^{1-1/p} \cdot \|x\|_p = k^{1-1/p}$. \square

3.4 Bounding Equation (3.1) for $b > 1$

The following estimate upper bounds the right hand side of (3.1) for any block $b \geq 1$, but we will use it eventually only for $b > 1$. For $b = 1$, we will need a separate estimate.

Lemma 3.7. *For every integer $1 \leq b \leq \lceil k/\ell \rceil$, we have*

$$\sum_{i=1}^m \sum_{j \in B_b} \left(|A_{ij} x_j| \cdot \left| \sum_{j'=j+1}^k A_{ij'} x_{j'} \right|^{p-1} \right) \leq ((3\delta dk)^{2-p} + (\delta d)^{2-p} \ell) \cdot \frac{S^{p-1}}{d^{1/p} \cdot ((b-1)\ell + 1)^{1/p}},$$

where S is defined in the statement of Lemma 3.6.

Proof. Let us partition the sum of interest into primary and secondary entries:

$$\begin{aligned} & \sum_{i=1}^m \sum_{j \in B_b} \left(|A_{ij} x_j| \cdot \left| \sum_{j'=j+1}^k A_{ij'} x_{j'} \right|^{p-1} \right) \\ & \leq \sum_{(i,j) \in L_b} \left(|A_{ij} x_j| \cdot \left| \sum_{j'=j+1}^k A_{ij'} x_{j'} \right|^{p-1} \right) + \sum_{(i,j) \in D_b} \left(|A_{ij} x_j| \cdot \left| \sum_{j'=j+1}^k A_{ij'} x_{j'} \right|^{p-1} \right) =: I + I'. \end{aligned} \quad (3.6)$$

Now, we upper bound I as follows:

$$\sum_{(i,j) \in L_b} \left(|A_{ij} x_j| \cdot \left| \sum_{j'=j+1}^k A_{ij'} x_{j'} \right|^{p-1} \right) \leq \frac{1}{d^{1/p} ((b-1)\ell + 1)^{1/p}} \cdot \sum_{(i,j) \in L_b} \left| \sum_{j'=j+1}^k A_{ij'} x_{j'} \right|^{p-1},$$

where the inequality follows from the fact that $|x_j| \leq \frac{1}{j^{1/p}}$ (since the coordinates of x are sorted in the decreasing order of their absolute values). We observe that we can apply Lemma 3.4 to the sum $\sum_{(i,j) \in L_b} \left| \sum_{j'=j+1}^k A_{ij'} x_{j'} \right|^{p-1}$, where the outer sum has at most $3\delta dk$ non-zero terms.¹¹ As a result, we have

$$\begin{aligned} \sum_{(i,j) \in L_b} \left(|A_{ij} x_j| \cdot \left| \sum_{j'=j+1}^k A_{ij'} x_{j'} \right|^{p-1} \right) & \leq \frac{(3\delta dk)^{2-p}}{d^{1/p} ((b-1)\ell + 1)^{1/p}} \cdot \left(\sum_{(i,j) \in L_b} \sum_{j'=j+1}^k |A_{ij'} x_{j'}| \right)^{p-1} \\ & \leq \frac{(3\delta dk)^{2-p} \cdot S^{p-1}}{d^{1/p} ((b-1)\ell + 1)^{1/p}} \end{aligned} \quad (3.7)$$

where the second inequality follows from Lemma 3.6, as $\sum_{(i,j) \in L_b} \sum_{j'=j+1}^k |A_{ij'} x_{j'}| = \sum_{(i,j) \in D_b \cup D'_b} A_{ij} |x_j|$.

¹¹This holds, since for every i, j, j' such that $(i, j) \in L_b$, $j' > j$ and $A_{ij'} \neq 0$ we have $(i, j') \in D_b \cup D'_b$ (see Figure 1). Due to Lemma 3.5 we have $|D_b \cup D'_b| \leq 3\delta dk$.

Next, we upper bound I' . For $i \in [m]$ we define $c_i := |\{j \mid (i, j) \in D_b\}| \leq \ell$. We have

$$\begin{aligned}
\sum_{(i,j) \in D_b} \left(|A_{ij}x_j| \cdot \left| \sum_{j'=j+1}^k A_{ij'}x_{j'} \right|^{p-1} \right) &\stackrel{\textcircled{1}}{\leq} \frac{1}{d^{1/p}((b-1)\ell+1)^{1/p}} \cdot \sum_{(i,j) \in D_b} \left| \sum_{j'=j+1}^k A_{ij'}x_{j'} \right|^{p-1} \\
&\stackrel{\textcircled{2}}{=} \frac{1}{d^{1/p}((b-1)\ell+1)^{1/p}} \cdot \sum_{i=1}^m c_i \left| \sum_{j'=j+1}^k A_{ij'}x_{j'} \right|^{p-1} \\
&\stackrel{\textcircled{3}}{\leq} \frac{\|c\|_{1/(2-p)}}{d^{1/p}((b-1)\ell+1)^{1/p}} \cdot \left(\sum_{(i,j) \in D_b} \sum_{j'=j+1}^k |A_{ij'}x_{j'}| \right)^{p-1} \\
&\stackrel{\textcircled{4}}{\leq} \frac{\|c\|_{1/(2-p)} \cdot S^{p-1}}{d^{1/p}((b-1)\ell+1)^{1/p}}.
\end{aligned}$$

Here, inequality $\textcircled{1}$ follows from the fact that $|x_j| \leq \frac{1}{j^{1/p}}$, equality $\textcircled{2}$ follows from the definition of c_i , inequality $\textcircled{3}$ follows from Lemma 3.4, and inequality $\textcircled{4}$ follows from Lemma 3.6.

Observe that for every $1 \leq q \leq \infty$ we have $\|c\|_q \leq \|c\|_\infty^{1-1/q} \|c\|_1^{1/q}$. From Lemma 3.5 we have $\|c\|_1 = |D_b| \leq \delta d \ell$, also by definition of c_i we have $\|c\|_\infty \leq \ell$. Overall, we obtain

$$\|c\|_{1/(2-p)} \leq \|c\|_\infty^{p-1} \cdot \|c\|_1^{2-p} \leq (\delta d)^{2-p} \cdot \ell. \quad (3.8)$$

We conclude by combining the upper bound on I' and (3.8) as follows:

$$\sum_{(i,j) \in D_b} \left(|A_{ij}x_j| \cdot \left| \sum_{j'=j+1}^k A_{ij'}x_{j'} \right|^{p-1} \right) \leq \frac{(\delta d)^{2-p} \ell \cdot S^{p-1}}{d^{1/p}((b-1)\ell+1)^{1/p}}. \quad (3.9)$$

Combining (3.6), (3.7) and (3.9), we get the desired inequality. \square

3.5 Bounding Equation (3.1) for $b = 1$

The following estimate upper bounds the right hand side of (3.1) for the block $b = 1$. It is tighter than that of Lemma 3.7.

Lemma 3.8. *For $b = 1$ one has*

$$\sum_{i=1}^m \sum_{j \in B_b} \left(|A_{ij}x_j| \cdot \left| \sum_{j'=j+1}^k A_{ij'}x_{j'} \right|^{p-1} \right) \leq O_\tau \left(d^{2-p} \cdot \frac{S^{p-1}}{d^{1/p}} + \delta \ell^{1-1/p} k^{(p-1)^2/p} \right),$$

where S is from the statement of Lemma 3.6.

Proof. We again decompose the sum according to the primary and secondary entries:

$$\begin{aligned}
&\sum_{i=1}^m \sum_{j \in B_b} \left(|A_{ij}x_j| \cdot \left| \sum_{j'=j+1}^k A_{ij'}x_{j'} \right|^{p-1} \right) \\
&\leq \sum_{(i,j) \in L_b} \left(|A_{ij}x_j| \cdot \left| \sum_{j'=j+1}^k A_{ij'}x_{j'} \right|^{p-1} \right) + \sum_{(i,j) \in D_b} \left(|A_{ij}x_j| \cdot \left| \sum_{j'=j+1}^k A_{ij'}x_{j'} \right|^{p-1} \right) =: I + I'. \quad (3.10)
\end{aligned}$$

First, let us upper bound I .

$$\begin{aligned}
\sum_{(i,j) \in L_b} \left(|A_{ij}x_j| \cdot \left| \sum_{j'=j+1}^k A_{ij'}x_{j'} \right|^{p-1} \right) &\stackrel{\textcircled{1}}{\leq} \left(\sum_{(i,j) \in L_b} |A_{ij}x_j|^{1/(2-p)} \right)^{2-p} \cdot \left(\sum_{(i,j) \in L_b} \sum_{j'=j+1}^k |A_{ij'}x_{j'}| \right)^{p-1} \\
&\stackrel{\textcircled{2}}{\leq} \left(\sum_{(i,j) \in L_b} (A_{ij} \cdot j^{-1/p})^{1/(2-p)} \right)^{2-p} \cdot S^{p-1} \\
&\stackrel{\textcircled{3}}{\leq} \frac{1}{d^{1/p}} \cdot \left| \sum_{j=1}^{\ell} d \cdot j^{-\frac{1}{p(2-p)}} \right|^{2-p} \cdot S^{p-1} \stackrel{\textcircled{4}}{\leq} \frac{S^{p-1}}{d^{1/p}} \cdot O_{\tau}(d^{2-p}).
\end{aligned} \tag{3.11}$$

Here, inequality $\textcircled{1}$ follows from Lemma 3.4. Inequality $\textcircled{2}$ follows from the fact that $|x_j| \leq \frac{1}{j^{1/p}}$, and Lemma 3.6 (since $\sum_{(i,j) \in L_b} \sum_{j'=j+1}^k |A_{ij'}x_{j'}| = \sum_{(i,j) \in D_b \cup D'_b} A_{ij}|x_j|$). Inequality $\textcircled{3}$ follows from the fact that there are at most d primary entries in the j -th column of the matrix for each $j \in B_1 = \{1, \dots, \ell\}$. Inequality $\textcircled{4}$ follows from the fact that $\sum_{j=1}^{\ell} j^{-\frac{1}{p(2-p)}} = O_{\tau}(1)$ when $\frac{1}{p(2-p)} \geq 1 + \Omega_{\tau}(1)$ (which is true because $1 + \tau \leq p \leq 2 - \tau$).

Next, let us upper bound I' . We note that

$$\begin{aligned}
\sum_{(i,j) \in D_b} \left(|A_{ij}x_j| \cdot \left| \sum_{j'=j+1}^k A_{ij'}x_{j'} \right|^{p-1} \right) &\stackrel{\textcircled{1}}{\leq} \left(\sum_{(i,j) \in D_b} |A_{ij}x_j| \right) \cdot \max_{(i,j) \in D_b} \left| \sum_{j'=j+1}^k A_{ij'}x_{j'} \right|^{p-1} \\
&\stackrel{\textcircled{2}}{\leq} \left(\sum_{(i,j) \in D_b} |A_{ij}x_j| \right) \cdot \left(\frac{\|x\|_1}{d^{1/p}} \right)^{p-1} \stackrel{\textcircled{3}}{\leq} \left(\sum_{(i,j) \in D_b} |A_{ij}x_j| \right) \cdot \left(\frac{k^{1-1/p}}{d^{1/p}} \right)^{p-1}.
\end{aligned}$$

Here, inequality $\textcircled{1}$ is obvious, inequality $\textcircled{2}$ follows from $|\sum_{j'=1}^k A_{ij'}x_{j'}| \leq \frac{1}{d^{1/p}}\|x\|_1$, and inequality $\textcircled{3}$ follows from $\|x\|_1 \leq k^{1-1/p}$. Since A expands B_1 , by [BGI⁺08, Lemma 9] we have¹²

$$\sum_{(i,j) \in D_b} |A_{ij}x_j| \leq \delta d^{1-1/p} \|x_{B_1}\|_1 \leq \delta d^{1-1/p} \cdot \sum_{t=1}^{\ell} t^{-1/p} \leq \delta d^{1-1/p} \cdot \int_0^{\ell} x^{-1/p} dx = \frac{\delta(d\ell)^{1-1/p}}{1-1/p},$$

where the second inequality follows from $|x_j| \leq j^{-1/p}$. In sum, we have

$$\sum_{(i,j) \in D_b} \left(|A_{ij}x_j| \cdot \left| \sum_{j'=j+1}^k A_{ij'}x_{j'} \right|^{p-1} \right) \leq O_{\tau} \left(\delta(d\ell)^{1-1/p} \cdot \left(\frac{k^{1-1/p}}{d^{1/p}} \right)^{p-1} \right) = O_{\tau} \left(\delta \ell^{1-1/p} k^{(p-1)^2/p} \right). \tag{3.12}$$

Finally, combining (3.10), (3.11) and (3.12), we get the desired inequality. \square

3.6 Proof of Theorem 3.1

Finally, we are ready to prove Theorem 3.1. We begin with a simple claim.

¹²The proof of this is similar to that of Lemma 3.6. In short, $\sum_{(i,j) \in D_b} |A_{ij}x_j| = d^{-1/p} \cdot \sum_{j=1}^{\ell} |x_j| \cdot |\{i \mid (i,j) \in D_b\}|$. Denoting by $a_j = |\{i \mid (i,j) \in D_b\}|$, we can rewrite this sum as $d^{-1/p} \cdot \sum_{j=1}^{\ell} a_j |x_j|$. Now, due to the expansion of A , we have $a_1 + \dots + a_t \leq \delta dt$ for every t ; on the other hand, $|x_j|$ is non-increasing as j increases. Overall, we conclude that this sum is maximized when $a_1 = \dots = a_t = \delta d$, and therefore, we obtain $\sum_{(i,j) \in D_b} |A_{ij}x_j| \leq \delta d^{1-1/p} \sum_{j=1}^{\ell} |x_j|$.

Claim 3.9. *One has*

$$\sum_{b=2}^{\lceil k/\ell \rceil} ((b-1)\ell + 1)^{-1/p} \leq O_\tau \left(\frac{k^{1-1/p}}{\ell} \right) .$$

Proof.

$$\sum_{b=2}^{\lceil k/\ell \rceil} ((b-1)\ell + 1)^{-1/p} \leq \int_1^{\lceil k/\ell \rceil} \frac{dx}{((x-1)\ell + 1)^{1/p}} \leq \frac{1}{\ell} \cdot \int_1^{2k} \frac{du}{u^{1/p}} \leq O_\tau \left(\frac{k^{1-1/p}}{\ell} \right) . \quad \square$$

Proof of Theorem 3.1. Combining (3.1), Lemma 3.7, Lemma 3.8, Claim 3.9, and that $S = 3\delta(dk)^{1-1/p}$, we get

$$\begin{aligned} & | \|Ax\|_p^p - 1 | \\ & \leq O(1) \cdot \sum_{b=1}^{\lceil k/\ell \rceil} \sum_{i=1}^m \sum_{j \in B_b} \left(|A_{ij}x_j| \cdot \left| \sum_{j'=j+1}^k A_{ij'}x_{j'} \right|^{p-1} \right) \quad (\text{using (3.1)}) \\ & \leq \sum_{b=2}^{\lceil k/\ell \rceil} O_\tau \left(((3\delta dk)^{2-p} + (\delta d)^{2-p}\ell) \cdot \frac{S^{p-1}}{d^{1/p} \cdot ((b-1)\ell + 1)^{1/p}} \right) + O_\tau \left(d^{2-p} \cdot \frac{S^{p-1}}{d^{1/p}} + \delta \ell^{1-1/p} k^{(p-1)^2/p} \right) \\ & \quad (\text{using Lemma 3.7 and Lemma 3.8}) \\ & \leq O_\tau \left(\left(\frac{k^{1-1/p}}{\ell} \cdot ((\delta dk)^{2-p} + (\delta d)^{2-p}\ell) + d^{2-p} \right) \cdot \frac{S^{p-1}}{d^{1/p}} + \delta \ell^{1-1/p} k^{(p-1)^2/p} \right) \quad (\text{using Claim 3.9}) \\ & = O_\tau \left(\left(\frac{k^{1-1/p}}{\ell} \cdot ((\delta dk)^{2-p} + (\delta d)^{2-p}\ell) + d^{2-p} \right) \cdot \delta^{p-1} \cdot d^{p-2} \cdot k^{(p-1)^2/p} + \delta \ell^{1-1/p} k^{(p-1)^2/p} \right) \\ & = O_\tau (\delta k \ell^{-1} + \delta k^{p-1} + \delta^{p-1} k^{(p-1)^2/p} + \delta \ell^{1-1/p} k^{(p-1)^2/p}) . \end{aligned}$$

We want this expression to be at most ε . For this, we can set $\ell = \Theta_\tau(k^{2-p}) \geq 1$ (note that we can do so because $p < 2$), and

$$\delta = \Theta_\tau \left(\min \left\{ \frac{\varepsilon}{k^{p-1}}, \frac{\varepsilon^{1/(p-1)}}{k^{(p-1)/p}} \right\} \right) .$$

Above, when deducing that $\delta \ell^{1-1/p} k^{(p-1)^2/p} \leq O(\varepsilon)$, we have used the fact that $\varepsilon < 1$.

Finally, from Lemma 3.2 we can choose $d = O(\frac{\log n}{\delta})$ and get the following number of rows:

$$m = O\left(\frac{dl}{\delta}\right) = O\left(\frac{\ell \cdot \log n}{\delta^2}\right) = O_\tau \left(\max \left\{ k^p \frac{\log n}{\varepsilon^2}, k^{4-2/p-p} \frac{\log n}{\varepsilon^{2/(p-1)}} \right\} \right) .$$

This finishes the proof of Theorem 3.1. \square

4 Dimension Lower Bounds

In this section, we prove dimension lower bounds for RIP- p matrices.

Theorem 4.1. *Let A be an $m \times n$ (k, D) -RIP- p matrix with distortion $D > 1$. Then,*

$$\begin{aligned} \text{If } 1 < p < 2, \quad & \text{either } m \geq \Omega\left(\frac{(2-p)n}{pD^2}\right)^{p/2} \quad \text{or } m \geq \Omega\left(\frac{k^p}{D^{2p/(2-p)}}\right) , \\ \text{If } p > 2, \quad & \text{either } m \geq \frac{n}{2k} \quad \text{or } m \geq \Omega\left(\frac{k^p}{D^{p^2/(p-2)}}\right) . \end{aligned}$$

4.1 Three Auxiliary Lemmas

We start with three auxiliary lemmas. The first one establishes bounds on the sum of p -th powers of the entries of A .

Lemma 4.2. *For any column $j \in [n]$, the following holds: $1 \leq \sum_{i=1}^m |A_{i,j}|^p \leq D^p$.*

Proof. $\sum_{i=1}^m |A_{i,j}|^p$ can be viewed as $\|Ae_j\|_p^p$. Now, for each $j \in [n]$, due to the (k, D) -RIP- p property, we have $1 \leq \|Ae_j\|_p^p \leq D^p$ completing the proof. \square

Next, for any $i \in [m]$ and $t \in [n]$, we denote by $b_{i,t}$ be the t -th largest absolute value in row i , that is, the t -th largest value among $|A_{i,1}|, |A_{i,2}|, \dots, |A_{i,n}|$. The following lemma establishes upper bounds on individual entries of A . Its proof relies on the RIP property for a k -sparse vector $x \in \{-1, 1\}^n$, chosen so that its entries ‘match’ the sign of the entries of A .

Lemma 4.3. *We have*

$$\max_{i \in [m]} b_{i,t} \leq \begin{cases} D \cdot t^{1/p-1}, & \text{if } t \leq k; \\ D \cdot k^{1/p-1}, & \text{if } t > k. \end{cases}$$

Proof. We first prove the lemma for any $t \leq k$. Consider any fixed row $i' \in [m]$. Let x be a t -sparse vector such that $x_j = \text{sgn}(A_{i',j})$ if $A_{i',j}$ is one of the $b_{i',1}, \dots, b_{i',t}$ and $x_j = 0$ otherwise. Then, the RIP- p property implies that

$$\|Ax\|_p^p = \sum_{i=1}^m \left| \sum_{j=1}^n A_{i,j} x_j \right|^p \leq D^p \cdot t.$$

In particular, since it is the sum over i of m non-negative terms, the above inequality also implies that for any specific row $i' \in [m]$:

$$\left| \sum_{j=1}^n A_{i',j} x_j \right|^p = \left(\sum_{t=1}^t |b_{i',t}| \right)^p \leq D^p \cdot t \implies \sum_{t=1}^t |b_{i',t}| \leq D \cdot t^{1/p}.$$

Since $|b_{i',t}|$ does not increase as t increases, we get $|b_{i',t}| \leq \frac{D \cdot t^{1/p}}{t} = D \cdot t^{1/p-1}$. This finishes the proof for $t = 1, 2, \dots, k$. For $t > k$, we have $|b_{i',t}| \leq |b_{i',k}| \leq D \cdot k^{1/p-1}$. \square

Our third lemma below establishes a lower (or upper) bound on the sum of squares of the entries of A . The proof of this lemma relies on the RIP property $\|Ax\|_p \approx \|x\|_p$ examined upon a *random* k -sparse vector x sampled from the uniform distribution over $\{-1, 1\}^k$.

Lemma 4.4. *If $1 < p \leq 2$ then $\sum_{i,j} A_{i,j}^2 \geq n \left(\frac{k}{m}\right)^{2/p-1}$; if $p \geq 2$ then $\sum_{i,j} A_{i,j}^2 \leq n D^2 \left(\frac{k}{m}\right)^{2/p-1}$.*

Proof. Let U be any set of k distinct indices in $[n]$ (i.e., columns). Let \mathcal{X} be the distribution of vectors $x \in \mathbb{R}^n$ such that $x_i = 0$ if $i \notin U$ and otherwise x_i is an independent random variable attaining values 1 and -1 with probability $1/2$ each. The (k, D) -RIP- p property implies that $k \leq \|Ax\|_p^p \leq D^p k$.

Let us now evaluate the following expectation:

$$\mathbb{E}_x[\|Ax\|_p^p] = \sum_{i=1}^m \mathbb{E}_x[|\langle A_i, x \rangle|^p] = \sum_{i=1}^m \mathbb{E}_x[(\langle A_i, x \rangle^2)^{p/2}] = \sum_{i=1}^m \sum_{x \in \mathcal{X}} \frac{(\langle A_i, x \rangle^2)^{p/2}}{2^k} \quad (4.1)$$

Comparing the $(p/2)$ -th power mean to the arithmetic mean, we have that if $p \leq 2$ then $\sum_{x \in \mathcal{X}} \frac{(\langle A_i, x \rangle^2)^{p/2}}{2^k} \leq (\frac{\sum_{x \in \mathcal{X}} \langle A_i, x \rangle^2}{2^k})^{p/2}$, and if $p \geq 2$ then $\sum_{x \in \mathcal{X}} \frac{(\langle A_i, x \rangle^2)^{p/2}}{2^k} \geq (\frac{\sum_{x \in \mathcal{X}} \langle A_i, x \rangle^2}{2^k})^{p/2}$. Finally, because of the way we have defined the distribution \mathcal{X} , we have $\sum_{x \in \mathcal{X}} \langle A_i, x \rangle^2 = 2^k \cdot \sum_{j \in U} A_{i,j}^2$.

Combining (4.1) with the above pieces for $1 < p \leq 2$ gives:

$$k \leq \mathbb{E}_x[\|Ax\|_p^p] \leq \sum_{i=1}^m (\sum_{j \in U} A_{i,j}^2)^{p/2}.$$

On the other hand, for $p \geq 2$, we get

$$\sum_{i=1}^m (\sum_{j \in U} A_{i,j}^2)^{p/2} \leq \mathbb{E}_x[\|Ax\|_p^p] \leq D^p k.$$

Let us first focus on the case of $1 < p \leq 2$. By again comparing the $(p/2)$ -th power mean to the arithmetic mean we can extend our inequality to

$$\frac{k}{m} \leq \frac{\sum_{i=1}^m (\sum_{j \in U} A_{i,j}^2)^{p/2}}{m} \leq \left(\frac{\sum_{i=1}^m \sum_{j \in U} A_{i,j}^2}{m} \right)^{p/2} \implies \sum_{i=1}^m \sum_{j \in U} A_{i,j}^2 \geq \frac{k^{2/p}}{m^{2/p-1}}. \quad (4.2)$$

Enumerating over all possible choices of indices U , we get the desired result:

$$\binom{n}{k} \cdot \left(\sum_{i=1}^m \sum_{j=1}^n A_{i,j}^2 \right) \cdot \frac{k}{n} = \sum_U \sum_{i=1}^m \sum_{j \in U} A_{i,j}^2 \geq \binom{n}{k} \cdot \frac{k^{2/p}}{m^{2/p-1}}.$$

If $p \geq 2$ then analogously to (4.2) we get the following inequality:

$$\left(\frac{\sum_{i=1}^m \sum_{j \in U} A_{i,j}^2}{m} \right)^{p/2} \leq \frac{\sum_{i=1}^m (\sum_{j \in U} A_{i,j}^2)^{p/2}}{m} \leq \frac{D^p k}{m} \implies \sum_{i=1}^m \sum_{j \in U} A_{i,j}^2 \leq \frac{D^2 k^{2/p}}{m^{2/p-1}},$$

and after enumerating over all possible sets of indices U gives: $\sum_{i=1}^m \sum_{j=1}^n A_{i,j}^2 \leq n D^2 \left(\frac{k}{m} \right)^{2/p-1}$. \square

4.2 Proof of Theorem 4.1

We first focus on the case of $1 < p < 2$. Using Lemma 4.3 we can evaluate

$$\sum_{i=1}^m \sum_{j=1}^k b_{i,j}^2 \leq \sum_{i=1}^m \sum_{j=1}^k \left(D \cdot j^{1/p-1} \right)^2 = m D^2 \sum_{j=1}^k j^{2/p-2} \leq m D^2 \int_{j=0}^k j^{2/p-2} dj \leq O\left(\frac{p}{2-p} m D^2 k^{2/p-1} \right) \quad (4.3)$$

and for the remaining terms:

$$\begin{aligned} \sum_{i=1}^m \sum_{j=k+1}^n b_{i,j}^2 &\stackrel{\textcircled{1}}{\leq} \sum_{i=1}^m \sum_{j=k+1}^n |b_{i,j}|^p \left(D \cdot k^{1/p-1} \right)^{2-p} \leq \left(\sum_{i=1}^m \sum_{j=1}^n |b_{i,j}|^p \right) \left(D \cdot k^{1/p-1} \right)^{2-p} \\ &\stackrel{\textcircled{2}}{=} \left(\sum_{i=1}^m \sum_{j=1}^n |A_{i,j}|^p \right) \left(D \cdot k^{1/p-1} \right)^{2-p} \stackrel{\textcircled{3}}{\leq} n \cdot D^p \left(D \cdot k^{1/p-1} \right)^{2-p}, \end{aligned} \quad (4.4)$$

where ① follows from Lemma 4.3, ② follows from the definition of $b_{i,j}$, and ③ follows from Lemma 4.2. Adding (4.3) and (4.4) gives:

$$\sum_{i,j} A_{i,j}^2 = \sum_{i=1}^m \sum_{j=1}^n b_{i,j}^2 \leq O\left(\frac{p}{2-p} m D^2 k^{2/p-1}\right) + n \cdot D^p \left(D \cdot k^{1/p-1}\right)^{2-p}$$

and using Lemma 4.4 we conclude that:

$$n \cdot \left(\frac{k}{m}\right)^{2/p-1} \leq O\left(\frac{p}{2-p} m D^2 k^{2/p-1}\right) + n \cdot D^p \left(D \cdot k^{1/p-1}\right)^{2-p}.$$

Therefore, either $n \cdot D^p (D \cdot k^{1/p-1})^{2-p}$ or $\frac{p}{2-p} m D^2 k^{2/p-1}$ must be at least $\Omega(n \cdot (\frac{k}{m})^{2/p-1})$. These two cases exactly correspond (after rearranging terms) to the desired inequalities.

(We remark here that when $p = 2$, the factor $(\frac{k}{m})^{2/p-1}$ on the left hand side becomes 1, and therefore no interesting lower bound on m can be deduced.)

Next, we focus on the case of $p > 2$. Let us compute again using Lemma 4.3:

$$\sum_{i=1}^m \sum_{j=k+1}^n b_{i,j}^2 \geq \sum_{i=1}^m \sum_{j=k+1}^n |b_{i,j}|^p \left(D \cdot k^{1/p-1}\right)^{2-p} \geq \left(\sum_{i=1}^m \sum_{j=k+1}^n |b_{i,j}|^p\right) \left(D \cdot k^{1/p-1}\right)^{2-p}.$$

Now, recall that the entries $\{b_{i,j}\}_{i,j}$ are by definition renamed from the entries of A , so the summation $\sum_{i=1}^m \sum_{j=k+1}^n |b_{i,j}|^p$ is missing precisely km entries from A . Therefore, this sum contains the p -th powers of all of the entries from at least $n - mk$ full columns of A , which is at least $n - mk$ (since any full column j of A , by Lemma 4.2, has its p -th power summing up to at least 1). Plugging this into the above inequality we get:

$$\sum_{i=1}^m \sum_{j=k+1}^n b_{i,j}^2 \geq (n - km) \left(D \cdot k^{1/p-1}\right)^{2-p}.$$

On the other hand,

$$\sum_{i,j} A_{i,j}^2 = \sum_{i=1}^m \sum_{j=1}^n b_{i,j}^2 \geq \sum_{i=1}^m \sum_{j=k+1}^n b_{i,j}^2 \geq (n - km) \left(D \cdot k^{1/p-1}\right)^{2-p},$$

and using Lemma 4.4 we conclude that

$$n D^2 \cdot \left(\frac{k}{m}\right)^{2/p-1} \geq (n - km) \left(D \cdot k^{1/p-1}\right)^{2-p}.$$

Now, we either have $m \geq \frac{n}{2k}$ or

$$\begin{aligned} D^2 \cdot \left(\frac{k}{m}\right)^{2/p-1} \geq \Omega\left(\left(D \cdot k^{1/p-1}\right)^{2-p}\right) &\implies \left(\frac{m}{k}\right)^{(p-2)/p} \geq \Omega\left(\frac{k^{(p-1)(p-2)/p}}{D^p}\right) \\ &\implies m \geq \Omega\left(\frac{k^p}{D^{p^2/(p-2)}}\right). \end{aligned}$$

Again, we emphasize that we used the strict inequality $p > 2$ in the above implication. \square

5 Column Sparsity Lower Bound

Below we provide a simple lower bound of $\Omega(k^{p-1})$ on the column sparsity of RIP- p matrices. The proof is a simple extension of an argument from [Cha10]. We remark that we are aware of an alternative proof of a slightly stronger lower bound that extends the argument of Nelson and Nguyễn [NN13], but since the better bound does not seem to be optimal, and the argument is much more complicated, we decided not to include its proof here.

Theorem 5.1. *Let A be an $m \times n$ matrix with (k, D) -RIP- p property and column sparsity d . Then, either $m > n/k$, or $d \geq k^{p-1}/D^p$.*

Proof. Assume that $m \leq n/k$. Since for every basis vector $e_j \in \mathbb{R}^n$ we have $\|Ae_j\|_p \geq 1$, it implies that for every column of A there is an entry with absolute value at least $d^{-1/p}$. Thus, there exists a row with at least $n/m \geq k$ such entries. Without loss of generality, let us assume that this is the first row, and the entries are located in columns from 1 to k . There exists a k -sparse vector x such that

- for every $1 \leq j \leq k$ we have $x_j = \text{sgn}(A_{1j}) \in \{-1, 1\}$;
- for every $j > k$ we have $x_j = 0$;
- the first coordinate of the vector Ax is at least $\frac{k}{d^{1/p}}$.

By the RIP property, we have $\frac{k}{d^{1/p}} \leq \|Ax\|_p \leq D \cdot \|x\|_p = D \cdot k^{1/p}$. Thus, $d \geq k^{p-1}/D^p$. \square

Acknowledgments

We thank Piotr Indyk for encouraging us to work on this project and for many valuable conversations. We are grateful to Piotr Indyk and Ronitt Rubinfeld for teaching “Sublinear Algorithms”, where parts of this work appeared as a final project. We thank Artūrs Bačkurs, Chinmay Hegde, Gautam Kamath, Sepideh Mahabadi, Jelani Nelson, Huy Nguyễn, Eric Price and Ludwig Schmidt for useful conversations and feedback. Thanks to Leonid Boytsov for pointing us to [Nag69a, Nag69b]. We are grateful to anonymous referees for pointing out some relevant literature. The first author is partly supported by a Simons Graduate Student Award under grant no. 284059.

References

- [AGMS14] Zeyuan Allen-Zhu, Rati Gelashvili, Silvio Micali, and Nir Shavit. Johnson-Lindenstrauss Compression with Neuroscience-Based Constraints. *ArXiv e-prints*, abs/1411.5383, November 2014. Also appeared in the Proceedings of the National Academy of Sciences of the USA, vol 111, no 47.
- [BDDW08] Richard Baraniuk, Mark Davenport, Ronald DeVore, and Michael Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3):253–263, 2008.
- [BGI⁺08] Radu Berinde, Anna C. Gilbert, Piotr Indyk, Howard Karloff, and Martin J. Strauss. Combining geometry and combinatorics: A unified approach to sparse signal recovery. In *Proceedings of the 46th Annual Allerton Conference on Communication, Control, and Computing (Allerton ’2008)*, pages 798–805, 2008.

- [BMRV02] Harry Buhrman, Peter Bro Miltersen, Jaikumar Radhakrishnan, and Srinivasan Venkatesh. Are bitvectors optimal? *SIAM Journal on Computing*, 31(6):1723–1744, 2002.
- [Can08] Emmanuel J. Candès. The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathématique*, 346(9–10):589–592, 2008.
- [Cha10] Venkat B. Chandar. *Sparse Graph Codes for Compression, Sensing, and Secrecy*. PhD thesis, Massachusetts Institute of Technology, 2010.
- [CRT06] Emmanuel Candès, Justin Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006.
- [CT05] Emmanuel Candès and Terence Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, 2005.
- [DIPW10] Khanh Do Ba, Piotr Indyk, Eric Price, and David P. Woodruff. Lower bounds for sparse recovery. In *Proceedings of the 21st Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '2010)*, pages 1190–1197, 2010.
- [Don06] David L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.
- [Efr65] Bradley Efron. Increasing properties of Pólya frequency functions. *Annals of Mathematical Statistics*, 36(1):272–279, 1965.
- [GI10] Anna C. Gilbert and Piotr Indyk. Sparse recovery using sparse matrices. *Proceedings of IEEE*, 98(6):937–947, 2010.
- [GSTV07] Anna C. Gilbert, Martin J. Strauss, Joel A. Tropp, and Roman Vershynin. One sketch for all: fast algorithms for compressed sensing. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing (STOC 2007)*, pages 237–246, 2007.
- [IR13] Piotr Indyk and Ilya Razenshteyn. On model-based RIP-1 matrices. In *Proceedings of the 40th International Colloquium on Automata, Languages, and Programming (ICALP '2013)*, pages 564–575, 2013.
- [JDP83] Kumar Joag-Dev and Frank Proschan. Negative association of random variables with applications. *Annals of Statistics*, 11(1):286–295, 1983.
- [KBG⁺10] Raghunandan M. Kainkaryam, Angela Bruex, Anna C. Gilbert, John Schiefelbein, and Peter J. Woolf. poolMC: Smart pooling of mRNA samples in microarray experiments. *BMC Bioinformatics*, 11(299), 2010.
- [Lat97] Rafał Łatała. Estimation of moments of sums of independent real random variables. *Annals of Probability*, 25(3):1502–1513, 1997.
- [LMN05] James R. Lee, Manor Mendel, and Assaf Naor. Metric structures in L_1 : dimension, snowflakes, and average distortion. *European Journal of Combinatorics*, 26(8):1180–1190, 2005.
- [Mut05] S. Muthukrishnan. *Data Streams: Algorithms and Applications*, volume 1 of *Foundations and Trends in Theoretical Computer Science*. 2005.

- [Nac10] Mergen Nachin. Lower bounds on the column sparsity of sparse recovery matrices. undergraduate thesis, MIT, 2010.
- [Nag69a] A.V. Nagaev. Integral limit theorems taking large deviations into account when Cramér’s condition does not hold. I. *Theory of Probability and Its Applications*, 14(1):51–64, 1969.
- [Nag69b] A.V. Nagaev. Integral limit theorems taking large deviations into account when Cramér’s condition does not hold. II. *Theory of Probability and Its Applications*, 14(2):193–208, 1969.
- [NN13] Jelani Nelson and Huy L. Nguyễn. Sparsity lower bounds for dimensionality reducing maps. In *Proceedings of the 45th ACM Symposium on the Theory of Computing (STOC ’2013)*, pages 101–110, 2013.
- [PS97] Alessandro Panconesi and Aravind Srinivasan. Randomized distributed edge coloring via an extension of the Chernoff-Hoeffding bounds. *SIAM Journal on Computing*, 26(2):350–368, 1997.
- [Rau10] Holger Rauhut. Compressive sensing and structured random matrices. *Theoretical foundations and numerical methods for sparse recovery*, 9:1–92, 2010.

A RIP Matrices and Stable Sparse Recovery

In this section we extend the main result from [CRT06] to the case of the general ℓ_p norms. Namely, we show that RIP- p matrices for $p > 1$ give rise to the polynomial-time stable sparse recovery with ℓ_p/ℓ_1 guarantee and approximation factors $C_1 = O(k^{-1+1/p})$ and $C_2 = O(1)$.

Suppose that we are given a sketch $y = Ax + e \in \mathbb{R}^m$ for a signal $x \in \mathbb{R}^n$, where $A \in \mathbb{R}^{m \times n}$, and $\|e\|_p \leq \varepsilon$. Our goal is to recover from y a good approximation \hat{x} to x . One of the standard ways to accomplish this is to solve the following ℓ_1 -minimization convex program:

$$\min_{\hat{x} \in \mathbb{R}^n} \|\hat{x}\|_1 \quad \text{such that} \quad \|A\hat{x} - y\|_p \leq \varepsilon . \quad (\text{A.1})$$

Let $S \subseteq [n]$ be the set of k largest (in absolute value) coordinates of x , and $h \stackrel{\text{def}}{=} \hat{x} - x$ be the error vector. For a parameter $\alpha > 0$ to be chosen later, we consider the following partition of $[n] \setminus S$: let $T_0 \subseteq [n] \setminus S$ be the set of αk largest (in absolute value) coordinates of h , let $T_1 \subseteq [n] \setminus (S \cup T_0)$ be the set of αk next largest coordinates, and so on. We state and prove some simple claims first that are true for *every* measurement matrix A .

Claim A.1.

$$\|Ah\|_p \leq 2\varepsilon$$

Proof.

$$\|Ah\|_p = \|A\hat{x} - Ax\|_p \leq \|A\hat{x} - y\|_p + \|Ax - y\|_p \leq 2\varepsilon ,$$

since \hat{x} is a feasible solution of (A.1), and $\|Ax - y\|_p = \|e\|_p \leq \varepsilon$. □

Claim A.2. *We have*

$$\|h_{\bar{S}}\|_1 \leq \|h_S\|_1 + 2\|x_{\bar{S}}\|_1 .$$

Proof. Since x is a feasible solution for (A.1), we have

$$\|x_S\|_1 + \|x_{\bar{S}}\|_1 = \|x\|_1 \geq \|\hat{x}\|_1 = \|x + h\|_1 \geq \|x_S\|_1 - \|h_S\|_1 + \|h_{\bar{S}}\|_1 - \|x_{\bar{S}}\|_1 . \quad \square$$

Claim A.3. For every $1 \leq p \leq \infty$ we have

$$\sum_{i \geq 1} \|h_{T_i}\|_p \leq \frac{1}{\alpha^{1-1/p}} \cdot \left(\|h_S\|_p + \frac{2\|x_{\bar{S}}\|_1}{k^{1-1/p}} \right) .$$

Proof. For every $i \geq 2$ we have $\|h_{T_i}\|_\infty \leq \|h_{T_{i-1}}\|_1/(\alpha k)$ by the definition of T_i , which implies $\|h_{T_i}\|_p \leq (\alpha k \cdot \|h_{T_i}\|_\infty^p)^{1/p} \leq \|h_{T_{i-1}}\|_1/(\alpha k)^{1-1/p}$. Hence

$$\begin{aligned} \sum_{i \geq 1} \|h_{T_i}\|_p &\leq \frac{1}{(\alpha k)^{1-1/p}} \cdot \sum_{i \geq 0} \|h_{T_i}\|_1 = \frac{\|h_{\bar{S}}\|_1}{(\alpha k)^{1-1/p}} \\ &\leq \frac{\|h_S\|_1 + 2\|x_{\bar{S}}\|_1}{(\alpha k)^{1-1/p}} \leq \frac{1}{\alpha^{1-1/p}} \cdot \left(\|h_S\|_p + \frac{2\|x_{\bar{S}}\|_1}{k^{1-1/p}} \right) , \end{aligned}$$

where the second inequality follows from Claim A.2 and the third inequality follows from the relation between ℓ_1 and ℓ_p norms, that is, $\|h_S\|_1 \leq k^{1-1/p} \cdot \|h_S\|_p$. \square

Claim A.4. For every $1 \leq p \leq \infty$ we have

$$\|h_{S \cup T_0}\|_p \leq \frac{1}{\alpha^{1-1/p}} \cdot \left(\|h_S\|_p + \frac{2\|x_{\bar{S}}\|_1}{k^{1-1/p}} \right) .$$

Proof.

$$\|h_{S \cup T_0}\|_p = \left\| \sum_{i \geq 1} h_{T_i} \right\|_p \leq \sum_{i \geq 1} \|h_{T_i}\|_p \leq \frac{1}{\alpha^{1-1/p}} \cdot \left(\|h_S\|_p + \frac{2\|x_{\bar{S}}\|_1}{k^{1-1/p}} \right) ,$$

where the last inequality follows from Claim A.3. \square

A.1 RIP- p matrices implies ℓ_p/ℓ_1 recovery

Here we prove that if A is a matrix with RIP- p property, then the ℓ_1 -minimization in (A.1) recovers a vector that is close enough to x . We begin with an auxiliary estimate.

Lemma A.5. If A is an $((\alpha + 1)k, D)$ -RIP- p matrix for $p > 1$ and $1 < D < \alpha^{1-1/p}$, then

$$\|h_{S \cup T_0}\|_p \leq \frac{2D}{\alpha^{1-1/p} - D} \cdot \frac{\|x_{\bar{S}}\|_1}{k^{1-1/p}} + \frac{2\alpha^{1-1/p}}{\alpha^{1-1/p} - D} \cdot \varepsilon .$$

Proof.

$$\begin{aligned} 2\varepsilon &\stackrel{\textcircled{1}}{\geq} \|Ah\|_p \geq \|Ah_{S \cup T_0}\|_p - \sum_{i \geq 1} \|Ah_{T_i}\|_p \\ &\stackrel{\textcircled{2}}{\geq} \|h_{S \cup T_0}\|_p - D \cdot \sum_{i \geq 1} \|h_{T_i}\|_p \\ &\stackrel{\textcircled{3}}{\geq} \|h_{S \cup T_0}\|_p - \frac{D}{\alpha^{1-1/p}} \cdot \left(\|h_S\|_p + \frac{2\|x_{\bar{S}}\|_1}{k^{1-1/p}} \right) \geq \left(1 - \frac{D}{\alpha^{1-1/p}} \right) \cdot \|h_{S \cup T_0}\|_p - \frac{2D \cdot \|x_{\bar{S}}\|_1}{(\alpha k)^{1-1/p}} , \end{aligned}$$

where the inequality $\textcircled{1}$ is due to Claim A.1 both x and \hat{x} are feasible for (A.1), inequality $\textcircled{2}$ holds since A satisfies the RIP- p property and inequality $\textcircled{3}$ is due to Claim A.3. \square

Now we are ready to extend the result from [CRT06]. We prove that if a measurement matrix A has RIP- p property for $p > 1$, then one can perform the stable sparse recovery with the ℓ_p/ℓ_1 guarantee via ℓ_1 -minimization.

Theorem A.6. *For every $D > 1$, if A is a $((4D)^{p/(p-1)}k, D)$ -RIP- p matrix for some $p > 1$, then*

$$\|h\|_p \leq \frac{O(1)}{k^{1-1/p}} \cdot \|x_{\bar{S}}\|_1 + O(\varepsilon).$$

Proof. Setting $\alpha = (2D)^{p/(p-1)} > 2$, we have $(4D)^{p/(p-1)}k \geq 2^{p/(p-1)} \cdot \alpha \cdot k > (\alpha + 1)k$ and therefore the assumptions in Lemma A.5 hold. We proceed as follows.

$$\begin{aligned} \|h\|_p &\leq \|h_{S \cup T_0}\|_p + \|h_{\bar{S} \cup T_0}\|_p \stackrel{\textcircled{1}}{\leq} \|h_{S \cup T_0}\|_p + \frac{1}{\alpha^{1-1/p}} \cdot \left(\|h_S\|_p + \frac{2\|x_{\bar{S}}\|_1}{k^{1-1/p}} \right) \\ &\leq \left(1 + \frac{1}{\alpha^{1-1/p}} \right) \cdot \|h_{S \cup T_0}\|_p + \frac{2\|x_{\bar{S}}\|_1}{(\alpha k)^{1-1/p}} \\ &\stackrel{\textcircled{2}}{\leq} \left(1 + \frac{1}{\alpha^{1-1/p}} \right) \cdot \left(\frac{2D}{\alpha^{1-1/p} - D} \cdot \frac{\|x_{\bar{S}}\|_1}{k^{1-1/p}} + \frac{2\alpha^{1-1/p}}{\alpha^{1-1/p} - D} \cdot \varepsilon \right) + \frac{2\|x_{\bar{S}}\|_1}{(\alpha k)^{1-1/p}} \\ &\stackrel{\textcircled{3}}{\leq} \frac{O(1)}{k^{1-1/p}} \cdot \|x_{\bar{S}}\|_1 + O(\varepsilon). \end{aligned}$$

Above, inequality $\textcircled{1}$ follows from Claim A.4, inequality $\textcircled{2}$ follows from Lemma A.5 and the last inequality $\textcircled{3}$ holds because $\alpha^{1-1/p} = 2D$. \square

A.2 ℓ_p/ℓ_1 recovery implies RIP- p matrices

Here we present a simple argument that any matrix A with $\|A\|_p \leq 1$ that allows stable sparse recovery with the ℓ_p/ℓ_1 guarantee (with arbitrarily large C_1) *must be* (k, C_2) -RIP- p . First, observe that the recovery procedure must map $0 \in \mathbb{R}^m$ to $0 \in \mathbb{R}^n$, as long as C_1 is finite. Second, let $x \in \mathbb{R}^n$ be any k -sparse signal, and consider a sketch $y = Ax + e$, where $e = -Ax$ (thus, $y = 0$). Since we must recover $0 \in \mathbb{R}^m$ to $0 \in \mathbb{R}^n$, one has from (1.1)

$$\|x\|_p \leq C_2 \cdot \|e\|_p = C_2 \cdot \|Ax\|_p.$$

Combining this inequality with $\|Ax\|_p \leq \|x\|_p$ (which follows from $\|A\|_p \leq 1$), we obtain the result.